The Journal of

Economic Perspectives

A journal of the American Economic Association

Winter 2012

The Journal of

Economic Perspectives

A journal of the American Economic Association

Editorial offices:
Journal of Economic Perspectives
American Economic Association Publications
2403 Sidney St., #260
Pittsburgh, PA 15203
e-mail: <jep@jepjournal.org>

The *Journal of Economic Perspectives* gratefully acknowledges the support of Macalester College.
Registered in the U.S. Patent and Trademark Office (®).

Composed by American Economic Association Publications, Pittsburgh, Pennsylvania, U.S.A.
Printed by R. R. Donnelley Company, Jefferson City, Missouri, 65109, U.S.A.
No responsibility for the views expressed by the authors in this journal is assumed by the editors or by the American Economic Association.
THE JOURNAL OF ECONOMIC PERSPECTIVES (ISSN 0895-3309), Winter 2012, Vol. 26, No. 1. The *JEP* is published quarterly (February, May, August, November) by the American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203-2418. Annual dues for regular membership are $20.00, $30.00, or $40.00 depending on income; for an additional $15.00, you can receive this journal in print. Further information on subscriptions and the American Economic Association can be found at <http://www.vanderbilt.edu/AEA/>. Periodicals postage paid at Nashville, TN, and at additional mailing offices.
POSTMASTER: Send address changes to the *Journal of Economic Perspectives,* 2014 Broadway, Suite 305, Nashville, TN 37203. Printed in the U.S.A.

The Journal of
Economic Perspectives

Contents

Volume 26 • Number 1 • Winter 2012

Statement of Purpose

The *Journal of Economic Perspectives* attempts to fill a gap between the general interest press and most other academic economics journals. The journal aims to publish articles that will serve several goals: to synthesize and integrate lessons learned from active lines of economic research; to provide economic analysis of public policy issues; to encourage cross-fertilization of ideas among the fields of economics; to offer readers an accessible source for state-of-the-art economic thinking; to suggest directions for future research; to provide insights and readings for classroom use; and to address issues relating to the economics profession. Articles appearing in the journal are normally solicited by the editors and associate editors. Proposals for topics and authors should be directed to the journal office, at the address inside the front cover.

Policy on Data Availability

It is the policy of the *Journal of Economic Perspectives* to publish papers only if the data used in the analysis are clearly and precisely documented and are readily available to any researcher for purposes of replication. Details of the computations sufficient to permit replication must be provided. The Editor should be notified at the time of submission if the data used in a paper are proprietary or if, for some other reason, the above requirements cannot be met.

Policy on Disclosure

Authors of articles appearing in the *Journal of Economic Perspectives* are expected to disclose any potential conflicts of interest that may arise from their consulting activities, financial interests, or other nonacademic activities.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 3–28

Is There an Energy Efficiency Gap?

Hunt Allcott and Michael Greenstone

Many analysts of the energy industry have long believed that energy efficiency offers an enormous "win-win" opportunity: through aggressive energy conservation policies, we can both save money and reduce negative externalities associated with energy use. In 1979, Pulitzer Prize-winning author Daniel Yergin and the Harvard Business School Energy Project made an early version of this argument in the book *Energy Future*:

> If the United States were to make a serious commitment to conservation, it might well consume 30 to 40 percent less energy than it now does, and still enjoy the same or an even higher standard of living . . . Although some of the barriers are economic, they are in most cases institutional, political, and social. Overcoming them requires a government policy that champions conservation, that gives it a chance equal in the marketplace to that enjoyed by conventional sources of energy.

Thirty years later, consultancy McKinsey & Co. made a similar argument in its 2009 report, *Unlocking Energy Efficiency in the U.S. Economy*:

> Energy efficiency offers a vast, low-cost energy resource for the U.S. economy—but only if the nation can craft a comprehensive and innovative approach to

■ *Hunt Allcott is Assistant Professor of Economics, New York University, New York City, New York. Michael Greenstone is 3M Professor of Environmental Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts. Allcott is a Faculty Research Fellow and Greenstone is a Research Associate, both at the National Bureau of Economic Research, Cambridge, Massachusetts. Their e-mail addresses are ⟨hunt.allcott@nyu.edu⟩ and ⟨mgreenst@mit.edu⟩.*

doi=10.1257/jep.26.1.3

> unlock it. Significant and persistent barriers will need to be addressed at multiple levels to stimulate demand for energy efficiency and manage its delivery . . . If executed at scale, a holistic approach would yield gross energy savings worth more than $1.2 trillion, well above the $520 billion needed through 2020 for upfront investment in efficiency measures (not including program costs). Such a program is estimated to reduce end-use energy consumption in 2020 by 9.1 quadrillion BTUs, roughly 23 percent of projected demand, potentially abating up to 1.1 gigatons of greenhouse gases annually.

In economic language, the "win-win" argument is that government intervention to encourage energy efficiency can improve welfare for two reasons. First, the consumption of fossil fuels, which comprise the bulk of our current energy sources, causes externalities such as harm to human health, climate change, and constraints on the foreign policy objectives of energy-importing countries. Second, other forces such as imperfect information may cause consumers and firms not to undertake privately profitable investments in energy efficiency. These forces, which we refer to as "investment inefficiencies," would create what is popularly called an Energy Efficiency Gap: a wedge between the cost-minimizing level of energy efficiency and the level actually realized. Yergin, McKinsey & Co., and other analysts have argued that this gap represents a significant share of total energy use: in their view, the ground is littered with $20 bills that energy consumers have failed to pick up.

The energy efficiency policy debate often comingles these two types of market failures—energy use externalities and investment inefficiencies—causing imprecision in research questions and policy goals. In this paper, we distinguish between the two market failures and clarify their separate policy implications. If energy use externalities are the only market failure, it is well known that the social optimum is obtained with Pigouvian taxes or equivalent cap-and-trade programs that internalize these externalities into energy prices, and that substitute policies are often much less economically efficient. If investment inefficiencies also exist, the first-best policy is to address the inefficiency directly: for example, by providing information to imperfectly informed consumers. However, when these interventions are not fully effective and investment inefficiencies remain, policies that subsidize or mandate energy efficiency might increase welfare. The central economic question around energy efficiency is thus whether there are investment inefficiencies that a policy could correct—in other words, "Is there an Energy Efficiency Gap?"

We examine two classes of evidence on the existence and magnitude of investment inefficiencies that could cause the Energy Efficiency Gap. First, we examine choices made by consumers and firms, testing whether they fail to make investments that would increase utility or profits. Second, we focus on specific investment inefficiencies, testing for evidence consistent with each. After presenting the evidence, we discuss policy implications. Throughout the paper, we highlight how the economics of energy efficiency connects to important questions in other applied micro fields, including behavioral economics, industrial organization, and development microeconomics.

Three key conclusions arise. First, although there is a long literature assessing investment inefficiencies related to energy efficiency, this body of evidence frequently does not meet modern standards for credibility. A basic problem is that much of the evidence on the energy cost savings from energy efficiency comes from engineering analyses or observational studies that can suffer from a set of well-known biases. Furthermore, even if the energy cost savings were known, energy efficiency investments often have other unobserved costs and benefits, making it difficult to assess welfare effects. This problem is general to other economic applications: in order to argue that an agent is not maximizing an objective function, the analyst must credibly observe that objective function in full. We believe that there is great potential for a new body of credible empirical work in this area, both because the questions are so important and because there are significant unexploited opportunities for randomized controlled trials and quasi-experimental designs that have advanced knowledge in other domains.

Second, when one tallies up the available empirical evidence from different contexts, it is difficult to substantiate claims of a pervasive Energy Efficiency Gap. Some consumers appear to be imperfectly informed, and the evidence suggests that investment inefficiencies do cause an increase in energy use in various settings. However, the empirical magnitudes of the investment inefficiencies appear to be smaller, indeed substantially smaller, than the massive potential savings calculated in engineering analyses such as McKinsey & Co. (2009).

Third, because consumers are quite heterogeneous in the degree of their investment inefficiencies, it is crucial to design targeted policies. Subsidizing energy efficient durables, for example, changes relative prices for all consumers. While this policy will increase welfare for some consumers, such benefits must be traded off against distortions to consumers not subject to inefficiencies. Policy evaluations must therefore consider not just how much a policy increases energy efficiency, but what types of consumers are induced to become more energy efficient. Welfare gains will be larger from a policy that preferentially affects the decisions of consumers subject to investment inefficiencies.

Background Facts on Energy Demand

Overview of Energy Demand and Energy Efficiency

Table 1 presents the breakdown of total energy demand across the sectors of the U.S. economy. Much of our discussion focuses on household energy use and personal transportation instead of commercial and industrial energy use, because these are areas where inefficiencies of imperfect information might be more severe. In 2007, the average U.S. household spent $2,400 on gasoline for their autos and another $1,900 on natural gas, electricity, and heating oil (U.S. Bureau of Labor Statistics 2007). Of this latter figure, heating and cooling are the most significant end uses, which suggests that they may also be the areas where energy conservation could have the largest effect.

Table 1
U.S. Energy Use

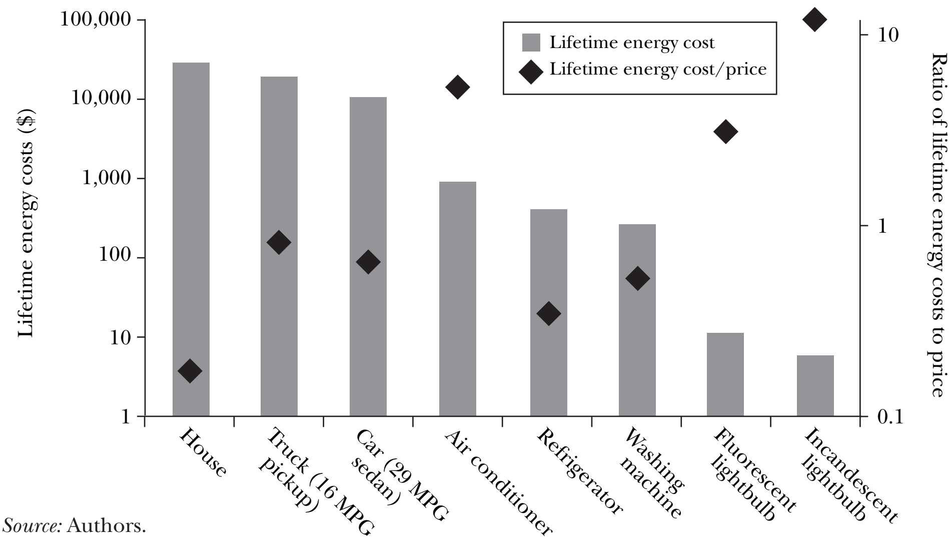
By sector (U.S. EIA 2011a)	
Commercial	19%
Industrial	30%
Transport	29%
Residential	22%
Residential categories (U.S. EIA 2005)	
Refrigerators	5%
Air conditioning	8%
Water heating	20%
Space heating	41%
Other appliances and lighting	26%

Source: Data are from U.S. Energy Information Administration (2005, 2011a).

The smaller the variance in energy costs across products relative to the total purchase price, the more likely it is that consumers will choose to remain imperfectly informed about, or inattentive to, these costs (Sallee 2011). Figure 1 shows the lifetime energy cost of a selection of energy-using durables, discounted at 6 percent over each good's typical lifetime, as well as the ratio of energy cost to the purchase price. For example, if gasoline costs $3 per gallon, lifetime gasoline costs are $19,000 for a typical pickup truck, or 83 percent of the purchase price, and $10,000 for a relatively energy efficient sedan, or about 66 percent of purchase price. Typical lifetime energy costs are five times greater than purchase prices for air conditioners and 12 times greater for incandescent light bulbs, but only about one-third of purchase price for a typical refrigerator.

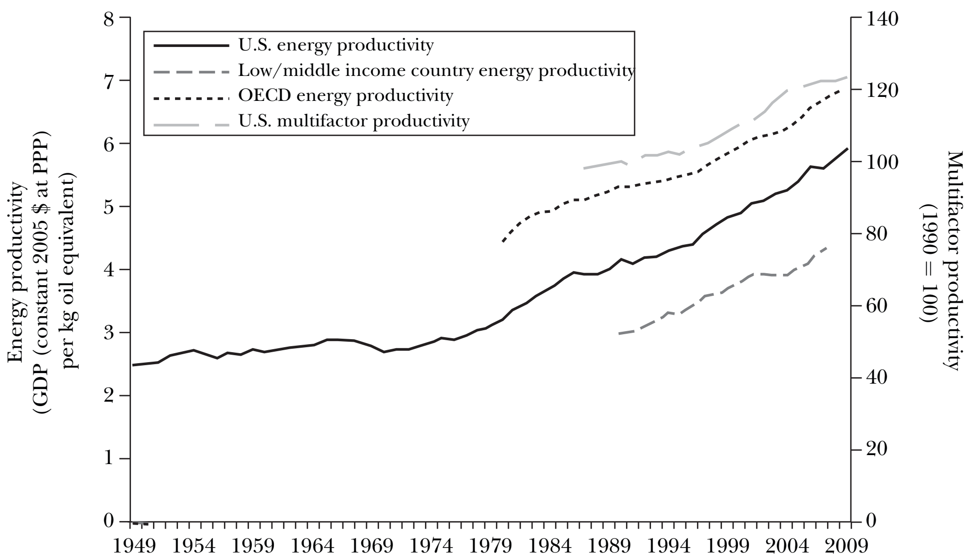
The most aggregate measure of energy efficiency is the ratio of GDP to total energy use, with different energy sources combined using common physical units. As shown in Figure 2, U.S. "energy productivity" per unit of GDP is 2.4 times higher than in 1949. Various factors drive this continual improvement, including compositional changes in the economy toward less-energy-intensive industries, energy efficiency policies, and other forces that drive total factor productivity growth. Energy prices also induce factor substitution and technical change: the figure suggests this effect, showing that the fastest improvements in energy productivity were in the 1970s and the most recent 15 years, both periods of relatively high energy prices. The figure also shows that U.S. energy productivity has grown faster than total factor productivity since the beginning of that data series in 1987, meaning that through some combination of directed technical change and factor substitution, the United States is economizing on energy faster than it is economizing on other factors. The U.S. economy is more energy intensive than other OECD countries, although it has improved more quickly since 1980, and less energy intensive than the set of low- and middle-income countries. In sum, the U.S. economy is progressively becoming less energy intensive, although this is uninformative about whether the United States is at or near the economically efficient level of energy efficiency.

Figure 1
Energy Costs for Durable Goods



Source: Authors.

Figure 2
Energy Productivity Trends



Sources: For U.S. energy input productivity 1949–1980, U.S. EIA Annual Energy Review (2011a), table 1.5 ⟨http://www.eia.gov/totalenergy/data/annual/showtext.cfm?t=ptb0105⟩; for U.S. multifactor productivity, U.S. Bureau of Labor Statistics, ⟨http://data.bls.gov/pdq/SurveyOutputServlet⟩; for other energy productivity, World Bank World Development Indicators, ⟨http://data.worldbank.org/indicator⟩.
Notes: PPP is "purchasing power parity." Multifactor productivity index equals 100 in 1990.

Table 2
Significant U.S. Energy Efficiency Policies

Name	*Year*	*Magnitude*
Corporate Average Fuel Economy Standards	1978–	$10 billion annual incemental cost from tightened 2012 rule (NHTSA 2010)
Federal Hybrid Vehicle Tax Credit	2006–2010	$426 million total annual credit (Sallee 2010)
Gas guzzler tax	1980–	$200 million annual revenues (Sallee 2010)
Federal appliance energy efficiency standards	1990–	$2.9 billion annual incremental cost (Gillingham, Newell, and Palmer 2006)
Residential and commercial building codes	1978–	
Electricity Demand-Side Management programs	1978–	$3.6 billion annual cost (US EIA 2010)
Weatherization Assistance Program (WAP)	1976–	$250 million annual cost (US DOE 2011a)
2009 Economic Stimulus	2009–2011	$17 billion total (U.S. DOE 2011b)
Additional WAP funding		$5 billion
Recovery Through Retrofit		$454 million
State Energy Program		$3.1 billion
Energy Efficiency and Conservation Block Grants		$3.2 billion
Home Energy Efficiency Tax Credits		$5.8 billion credit in 2009 (U.S. IRS 2011)
Residential and Commercial Building Initiative		$346 million
Energy Efficient Appliance Rebate Program		$300 million
Autos Cash for Clunkers		$5 billion

Source: Authors.

Energy Efficiency Policy in the United States

The United States has enacted a wide array of policies to encourage energy efficiency, many of which were originally promulgated during the energy crises of the 1970s. Table 2 presents the most significant of these policies, along with some measure of their annual costs. Auto industry policies include: Corporate Average Fuel Economy (CAFE) standards, which require that the new cars and trucks sold by each auto manufacturer meet a minimum average rating based on miles-per-gallon; tax credits of up to $3,400 for hybrid vehicle buyers; and "gas guzzler taxes" ranging from $1,000 to $7,700 on the sale of passenger cars with low fuel economy. There are a series of national-level minimum energy efficiency standards for household appliances, such as refrigerators, air conditioners, and washing machines. Additionally, many states have building codes that encourage energy efficiency by, for example, stipulating minimum amounts of required insulation. Furthermore, electricity bill surcharges fund billions of dollars of utility-managed "demand-side management" programs, which include subsidized residential and commercial energy audits, energy efficiency information provision, and subsidies for energy efficient appliances and other capital investments.

"Weatherization" is frequently used as a general term for a set of residential energy efficiency investments primarily including wall and attic insulation, improved heating, ventilation and air conditioning systems, and "air-sealing," which reduces the leakage of hot or cold outside air. Through the Weatherization Assistance Program, the federal government transfers $250 million annually to state agencies to weatherize approximately 100,000 low-income homes. Weatherization funding grew significantly due to the 2009 American Recovery and Reinvestment Act. In total, that legislation and related economic stimulus bills included $17 billion in energy efficiency spending, including non-low-income weatherization programs, automobile and appliance cash-for-clunkers programs with energy efficiency requirements on new models, and other grants to state programs.

In this paper, the phrase "energy efficiency policies" refers to this set of subsidies and standards that directly encourage investment in energy efficient capital stock but do not directly affect energy prices. Although gasoline taxes, cap-and-trade programs, or other policies that affect energy prices will of course also increase investment in energy efficient capital stock, these policies that act through energy prices are conceptually distinct in our policy analysis.

A Model of Investment in Energy Efficiency

The basic economics of energy efficiency are captured by a model in which an agent, either a profit-maximizing firm or utility-maximizing consumer, chooses between two different versions of an energy-using durable good such as an automobile, air conditioner, or light bulb.[1] This setup can also represent a choice of whether to improve the energy efficiency of an existing building, for example through weatherization. In the first period, the agent chooses and pays for capital investments. In the second period, the consumer uses the good and incurs energy costs.

The two different goods are denoted 0, for the energy inefficient baseline, and 1, for the energy efficient version. They have energy intensities e_0 and e_1, respectively, with $e_0 > e_1$. The energy efficient good has incremental upfront capital cost $c > 0$ and unobserved incremental opportunity cost or utility cost ξ. The variable ξ could either be positive (an unobserved cost) or negative (an unobserved benefit). The private cost of energy is p, and the risk-adjusted discount rate between the two periods is $r > 0$. The variable m represents an agent's taste for usage of the durable good; a high m reflects an air conditioner user in a hot climate or a car owner who drives a long way to work. The variable m is implicitly a function of energy prices: as energy prices

[1] The model presented here is an adaptation of the model in Allcott, Mullainathan, and Taubinsky (2011). It resembles a generalized Roy model. It abstracts away from factors which may be relevant in some settings, including the irreversibility of some energy efficiency investments and uncertainty over energy costs (Dixit and Pindyck 1994; Hassett and Metcalf 1993) and explicit models of imperfect information in the purchase or resale of the good.

rise, the cost of utilization increases, so utilization decreases. We index m_i to explicitly recognize that it varies across agents, although in practice ξ and p will also vary.

In the basic case, an agent's willingness-to-pay for the energy efficient good is the discounted energy cost savings net of unobserved costs. Agent i will choose the energy efficient good if and only if willingness to pay outweighs the incremental capital costs:

$$\frac{pm_i(e_0 - e_1)}{(1 + r)} - \xi > c.$$

To capture the essence of the Energy Efficiency Gap, we introduce the parameter γ, which is an implicit weight on the energy cost savings in the agent's decision. Now, the agent chooses the energy efficient good if and only if:

$$\frac{\gamma pm_i(e_0 - e_1)}{(1 + r)} - \xi > c.$$

For the purpose of determining the effects of subsidizing the energy efficient good, the γ parameter is a sufficient statistic for all investment inefficiencies. As we will discuss later in more detail, there are several distinct types of investment inefficiencies. First, agents may be unaware of, imperfectly informed about, or inattentive to energy cost savings. Second, agents may be themselves perfectly informed but unable to convey costlessly the energy intensity e_1 of an improved house or apartment they are selling or renting to others. Third, credit markets may be imperfect, meaning that agents may not have access to credit at the risk-adjusted discount rate r. [2] The γ parameter is conceptually related to what others have called an "implied discount rate," which is the discount rate that rationalizes the tradeoffs that agents make between upfront investment costs and future energy savings.

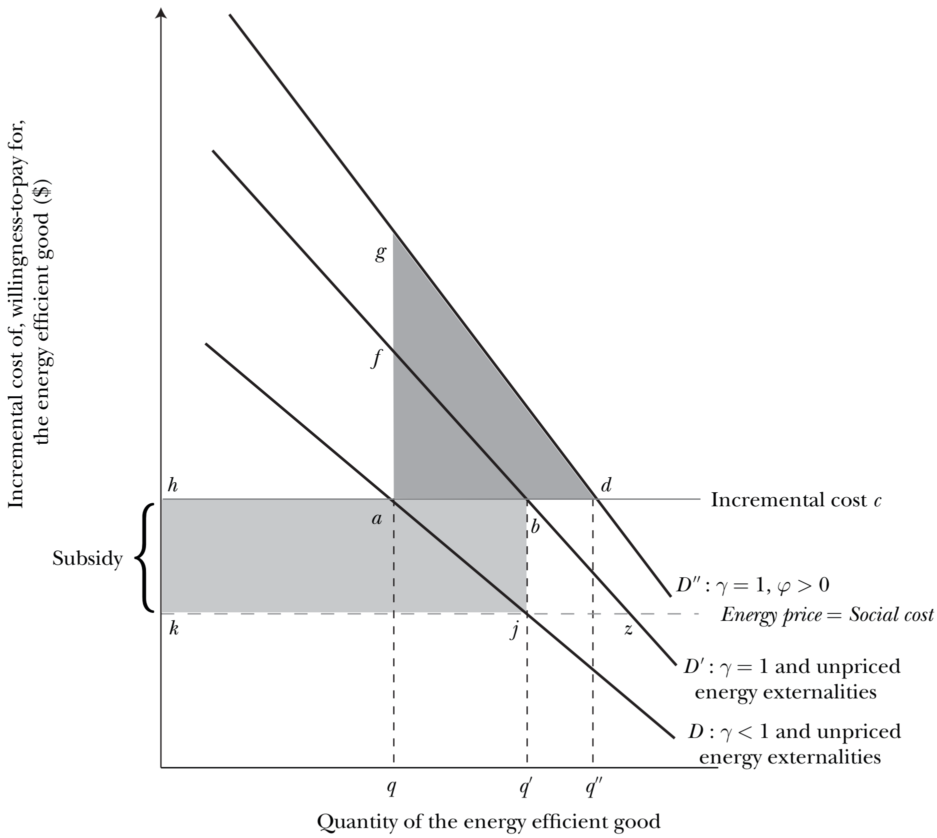
It is often asserted that $\gamma < 1$, meaning that investment inefficiencies cause agents to value discounted energy cost savings less than upfront costs. Notice that when this is the case, some agents do not choose the energy efficient good despite the fact that this would be profitable at current energy prices. Formally, asserting that there is an "Energy Efficiency Gap" is exactly equivalent to asserting that there are investment inefficiencies and $\gamma < 1$. Of course, in some settings it might be that $\gamma > 1$.

Other than the investment inefficiencies captured by γ, the additional element of the "win-win argument" is that there are additional social costs from energy use that are not internalized into energy prices. We denote this uninternalized externality by φ. In the social optimum, the agent adopts the energy efficient good if:

$$\frac{(p + \varphi)m_i(e_0 - e_1)}{(1 + r)} - \xi > c.$$

[2] Credit constraints are a frequently discussed investment inefficiency. Although we note the issue in theory, there is not much empirical evidence in the context of energy efficiency, so we will not discuss it further.

Figure 3
Demand for the Energy Efficient Good



The social optimum differs from the agent's choice in the previous equation for two reasons. First, the allocation accounts for the externality φ. Second, the allocation is not affected by investment inefficiencies, so $\gamma = 1$.

Figure 3 illustrates the three cases. The figure's horizontal axis represents the quantity of the energy efficient good that is purchased, while the vertical axis shows the incremental costs and benefits of purchasing that good. The height of a demand curve at each point reflects some individual agent's willingness-to-pay from the left-hand side of a corresponding equation above. The agents on the left side of the figure, with higher willingness-to-pay, tend to have high usage *m*, low unobserved cost ξ, and high energy price *p*.

The lowest demand curve, denoted *D*, reflects the case in the second equation with both investment inefficiencies ($\gamma < 1$) and uninternalized energy use externalities. In this case, the market equilibrium is at point *a*, the intersection of demand curve *D* with incremental cost *c*. Demand curve D' reflects the case in the first equation with no investment inefficiencies, but energy still priced below social cost. Demand curve D'' reflects the social optimum in the third equation, where there are no investment inefficiencies and energy prices include externality φ. Adding a Pigouvian tax on energy consumption (based on the energy source's pollution content) increases willingness-to-pay more for the consumers on the left of the figure with higher utilization, so demand curve D'' rotates clockwise relative to demand curve D'. The first-best equilibrium is point d, where D'' intersects incremental cost c.

From a policy perspective, it is crucial to distinguish the two types of market failures, energy use externalities and investment inefficiencies. The reason derives from the general principle that policies should address market failures as directly as possible. If there are no investment inefficiencies but energy prices are below social cost due to uninternalized energy use externalities, demand is represented by D'. This causes a distortion both in the purchase and in the utilization of energy-using durables: for example, consumers buy too many gas guzzlers and drive them too much. A Pigouvian tax of amount φ on energy (on gas, in the example) would give both the socially optimal quantity demanded (q'') of the energy efficient good and the socially optimal utilization. By contrast, as long as utilization is not fully price-inelastic, a subsidy for the energy efficient good does not achieve the first best. While this could move quantity demanded to q'', consumers would not face the true social cost of energy when deciding how much to use the good: consumers would buy the right number of gas guzzlers but still drive them too much.

Many investment inefficiencies, on the other hand, distort purchases but not utilization. If there are investment inefficiencies but no uninternalized energy use externalities, the optimal corrective policy affects purchases, but not utilization. For example, Allcott, Mullainathan, and Taubinsky (2011) show that when consumers have homogeneous $\gamma < 1$ and vary only in utilization m_i, the first-best policy involves a subsidy for the energy efficient good.[3] In Figure 3, that optimal subsidy would move quantity demanded from q to q'. Notice that an energy tax could potentially also correct the investment inefficiency, giving the same marginal consumer at q'. However, as long as utilization is not fully inelastic, an energy tax that gives price above social cost (to correct the investment inefficiency) would cause consumers to reduce utilization below the first-best level: consumers would buy the right number of gas guzzlers and then drive them too little.

Putting these arguments together, when there are distortions from both uninternalized energy use externalities and investment inefficiencies, the first-best policy involves both Pigouvian taxes on energy and a second mechanism to increase quantity demanded of the energy efficient good. This second mechanism may be a subsidy for the energy efficient good, although as we will discuss later in the paper, heterogeneity in the investment inefficiency γ makes subsidies potentially less desirable.

[3] Heutel (2011) obtains a comparable result using a different model of investment inefficiencies.

How can this framework be used for cost–benefit analysis? Consider first adding the subsidy in isolation, without any Pigouvian tax on energy. When there are investment inefficiencies, the original marginal consumer at quantity q gains amount *af* from being induced to buy the energy efficient good. In fact, there are allocative gains from inducing each of the consumers between q and q' to purchase the energy efficient good, as each of these consumers has benefits that are larger than incremental cost c. The total private welfare gains are illustrated by the triangle *abf*. If a Pigouvian tax on energy is added to this subsidy, then the total social welfare gain is illustrated by the triangle *adg*.

These benefits are then compared against the costs of the policy. A subsidy involves a transfer of public funds to consumers of amount *hbjk*, as illustrated by the shaded rectangle. If those funds could otherwise be used to lower labor taxes, the deadweight loss of these taxes would be included as a social cost, along with any other costs of administering the subsidy.[4] Similarly, an information program that moved demand from D to D' would also increase private welfare by *abf*, and this welfare gain would be traded off with the costs of implementation. For any policy, it will be an empirical question whether the costs exceed the benefits, and whether the net benefits are larger than alternative policies. This approach to assessing net welfare benefits is the appropriate test of whether energy efficiency policies are socially beneficial when Pigouvian taxes are also available to correct energy use externalities.

To summarize, this section has analyzed two forces that can cause behavior to differ from the social optimum: energy use externalities and investment inefficiencies. If there are energy use externalities but no investment inefficiencies, ideally only Pigouvian taxes would be used. If there are investment inefficiencies, energy efficiency policies such as subsidies for energy efficient capital stock might have benefits that outweigh their costs. If there are both investment inefficiencies and energy use externalities, then Pigouvian taxes should be used in combination with some welfare-improving energy efficiency policy. The central economic questions are thus whether there are investment inefficiencies, and if so, whether the benefits of a corrective policy outweigh its costs.

In the next section, we will examine choices by consumers and firms to adopt or not adopt energy efficient technologies and attempt to infer whether there is an Energy Efficiency Gap. When there are no investment inefficiencies, agents' choices are governed by the first equation above, and unobserved factors such as costs ξ or utilization m can be inferred from their decisions. Some analysts have relied heavily on this framework in explaining away an apparent Energy Efficiency Gap, with an argument along the lines that "agents are well-informed, so if they are not energy efficient, then it must be that the unobserved costs of energy efficiency are large." The analysis is more difficult when there might be investment inefficiencies. In that case, we now must know everything about agents' objective functions to estimate the size of γ.

[4] Analogously, a Pigouvian tax brings in public funds that can be used to lower labor taxes, which should be counted as an additional benefit (Bovenberg and Goulder 1996).

Three types of problems will pervade the analyses we review in the next section. First, factors that are difficult to observe or quantify, as denoted by ξ in our model above, will be potentially very relevant. Second, estimates of the net present value of energy cost savings are often questionable. Depending on the setting, this could be because the analyst does not know the change in energy intensity $(e_0 - e_1)$, the utilization m, or the appropriate discount rate r. Third, there is often substantial heterogeneity across consumers in utilization and unobserved costs, meaning that average returns for adopters might be uninformative about average returns for non-adopters or returns for the marginal adopter.

These empirical problems directly parallel other economic contexts. Consider, for example, the question of whether farmers in developing countries could profitably adopt agricultural technologies such as fertilizer and high-yielding variety seeds. These technologies have unobserved costs, such as increased labor inputs (Foster and Rosenzweig 2010). It is difficult to know the resulting increase in profits without randomized controlled trials, as in Duflo, Kremer, and Robinson (2011). Also, the substantial heterogeneity in costs and gross returns means that the fact that adopters have high returns does not imply that non-adopters are foregoing a profitable investment (Suri 2011).

Evidence on Returns to Energy Efficiency Investments

In this section, we analyze the evidence on whether consumers and firms leave profitable energy efficiency investments on the table. There are four categories of evidence: engineering estimates of returns to potential investments, empirical estimates of returns to observed investments, the cost effectiveness of energy conservation programs run by electric utilities, and estimated demand patterns for energy-using durables.

Engineering Estimates of Energy Conservation Cost Curves

While the McKinsey & Co. (2009) study quoted in our introduction has garnered substantial attention, it is preceded by a long literature that uses engineering cost estimates to construct "supply curves" for energy efficiency (for example, Meier, Wright, and Rosenfeld 1983; ACEEE 1989; Goldstein, Mowris, Davis, and Dolan 1990; Koomey et al. 1991; Brown, Levine, Romm, Rosenfeld, and Koomey 1998; National Academy of Sciences 1992; Rosenfeld, Atkinson, Koomey, Meier, Mowris, and Price 1993; Stoft 1995; Blumstein and Stoft 1995; Brown, Levine, Short, and Koomey 2001). The basic approach in such studies is to calculate the net present value of a set of possible energy efficiency investments given assumed capital costs, energy prices, investment horizons, and discount rates.

Across many studies from different industries and sectors, a common theme seems to emerge: large fractions of energy can be conserved at *negative* net cost. That is, the studies conclude that consumers and firms are failing to exploit a massive amount of profitable investment opportunities in energy efficiency. For example, a meta-analysis by Rosenfeld, Atkinson, Koomey, Meier, Mowris, and Price (1993) concludes that between 20 and 60 percent of total electricity use, depending on the study and the electricity cost assumption, can be conserved at negative cost. The McKinsey & Co. (2009) analysis quoted in our introduction suggests that 23 percent of U.S. nontransportation energy demand can be eliminated at negative cost. These engineering studies are a large part of the basis for the claims about the Energy Efficiency Gap.

However, it is difficult to take at face value the quantitative conclusions of the engineering analyses as they suffer from the empirical problems introduced in the previous section. First, engineering costs typically incorporate only upfront capital costs and omit opportunity costs or other unobserved factors (ξ in the model presented earlier). For example, Anderson and Newell (2004) analyze energy audits that the U.S. Department of Energy provides for free to small- and medium-sized enterprises. They find that nearly half of investments that engineering assessments showed would have short payback periods were not adopted due to unaccounted physical costs, risks, or opportunity costs, such as "lack of staff for analysis/implementation," "risk of inconvenience to personnel," or "suspected risk of problem with equipment."

Second, the engineering estimates of energy saved may be faulty. For example, in the context of home energy weatherization, Dubin, Miedema, and Chandran (1986), Nadel and Keating (1991), and others have documented that engineering estimates of energy savings can overstate true field returns, sometimes by a large amount. Even in the two decades since these studies, some engineering simulation models have still not been fully calibrated to approximate actual returns (Blasnik 2010).

Empirical Estimates of Returns on Investment

Another approach to measuring the Energy Efficiency Gap is to use empirical energy use data to estimate the average returns for the set of consumers that adopt an energy efficient technology. Most of the evidence in this category analyzes the costs and benefits of the Weatherization Assistance Program, which is intended to be both a transfer to low-income homeowners and an energy efficiency investment with positive net returns. The typical empirical analysis compares natural gas billing data in the first year after the weatherization work was done to the year before, using either a statistical correction for weather differences or a nonrandomly selected control group of low-income households. Schweitzer (2005) analyzes 38 separate empirical evaluations of weatherization projects from 19 states from between 1993 and 2005, reweighting them to reflect the observable characteristics of the national Weatherization Assistance Program. The average weatherization job costs $2,600 and reduces natural gas use by 20 to 25 percent, or about $260 per year.

As evidence on the Energy Efficiency Gap, such analyses again suffer from the problems introduced in the previous section. First, there are potentially substantial unobserved costs and benefits (the ξ in our model) from weatherization.

Weatherization takes time, and for most people it is not highly enjoyable: the process requires one or sometimes two home energy audits, a contractor appointment to carry out the work, and sometimes additional follow-up visits and paperwork. Some benefits are also difficult to quantify: for example, weatherization typically makes homes more comfortable and less drafty. Furthermore, weatherization reduces the cost of energy services such as warmer indoor temperatures on a cold winter day, and this cost reduction causes people to increase their utilization of these services. (In the energy literature, this is called the "rebound effect.") Measuring the change in energy use from weatherization without accounting for the utility gain from an increase in utilization of energy services understates the welfare benefits.

Second, the net present value of energy cost reductions is unknown. The empirical estimates are based on short-term analyses, and the persistence of returns over many years is rarely assessed. If the $260 annual savings from Schweitzer (2005) are assumed to have a lifetime of 10 years or less, then weatherization does not pay back the $2,600 cost at any positive discount rate. At lifetimes of 15 or 20 years, the discount rate that equates future discounted benefits with current costs (the internal rate of return) is 5.6 or 7.8 percent, respectively. Furthermore, all of the estimates are nonexperimental, and households that weatherize may also engage in other unobserved activities that affect energy use. This may be a larger concern with non-low-income weatherization programs, in which homeowners might be more likely to carry out renovations and energy efficiency work at the same time.

Third, the effects of weatherization on energy use are heterogeneous. For example, Metcalf and Hassett (1999) estimate the distribution of returns to attic insulation in the U.S. population using a weather-adjusted difference estimator with nationally representative panel data. The estimated median and mean returns on investment are on the order of 10 percent, and one-quarter of households had returns greater than 13.5 percent. This heterogeneity means that while estimates of average returns for adopters could in principle be meaningful in evaluating the costs and benefits of an existing program, a simple selection model like the one above would imply that the net returns for adopters overstate the net returns for non-adopters. On net, the available evidence seems inconsistent with significant investment inefficiencies in the context of weatherization.

Cost Effectiveness of Energy Conservation Programs

Many electric utilities run "demand-side management" programs, which largely consist of subsidies to households and firms to purchase energy efficient appliances, air conditioning and heating systems, and other equipment. If these programs can reduce energy use at less than the cost of energy, the argument goes, then there were investment inefficiencies, and the programs should be viewed as welfare-enhancing.

The simplest example of this approach is to divide the annual spending on these programs by estimates of electricity savings, as in Gillingham, Newell, and Palmer (2006). For 2009, U.S. electric utilities reported $2.255 billion in direct costs and 76.9 terawatt-hours of savings for demand-side management programs, according to the *2009 Electric Power Annual* (EIA 2010, tables 9.6 and 9.7). Dividing these two figures gives a cost effectiveness of 2.9 cents per kilowatt-hour (kWh).

Analyses such as these suffer from the same problems introduced in the previous section. First, the reported "costs" are typically costs to the utility, not including costs incurred by program participants, which may be almost as large (Nadel and Geller 1996; Joskow and Marron 1992; Eto, Kito, Shown, and Sonnenblick 1995; Friedrich, Eldridge, York, Witte, and Kushler 2009). Second, energy savings are estimated using observational data, and it is difficult to establish a credible counterfactual level of energy use in the absence of the program. The most rigorous solution is to use randomized controlled experiments to evaluate demand-side management programs. The feasibility of this approach is demonstrated by recent experimental evaluations of programs that send letters that compare a household's energy use to that of their neighbors and provide energy conservation tips (Allcott 2011b; Ayres, Raseman, and Shih 2009).

The most advanced estimate in this literature is by Arimura, Li, Newell, and Palmer (2011), whose point estimates indicate that between 1992 and 2006, demand-side management conserved electricity at a program cost of 5.0 and 6.1 cents per kilowatt hour, assuming discount rates of 5 and 7 percent, respectively. If one further assumes, based on the analyses in the paragraph above, that additional costs to consumers might be 70 percent of program costs, one concludes that demand-side management programs have reduced energy use at an average cost of $5.0 \times (1 + .70) = 8.5$ cents/kWh or $6.1 \times (1 + .70) = 10.4$ cents/kWh, again using 5 or 7 percent discount rates, respectively. Comparing the investment cost per kWh conserved to the national average electricity price of 9.1 cents/kWh, the investments that occurred because of demand-side management programs were barely profitable at a discount rate of 5 percent, and barely unprofitable at a discount rate of 7 percent.[5] Arimura et al. (2011) estimate that these programs reduced 1–2 percent of national electricity demand. Given that only a small percent of total electricity demand was reduced at nearly zero excess profits, this evidence on demand-side management energy conservation programs does not suggest a pervasive Energy Efficiency Gap.

Tradeoffs between Durable Goods

The final way of determining whether there are profitable returns to energy efficiency investments involves estimating consumer demand for household appliances or automobiles. This approach typically uses a discrete choice model to estimate utility function coefficients on purchase price and on the present discounted value of energy costs. The estimated coefficient on energy cost should be the same as the

[5] The text offers a back-of-the-envelope version of the more sophisticated calculation that should be done. One implicit assumption is that there are no consumers that are inframarginal to the demand-side management subsidies. If there are inframarginal consumers, then some of the program costs were in fact transfers, and the incremental investment costs induced by the demand-side management programs were smaller.

estimated coefficient on price: that is, consumers should be indifferent between spending a dollar in present value on energy and a dollar in present value on purchase price. If the analyst's assumptions about discount rates, product utilization, and energy prices are correct, the ratio of these two coefficients is the γ in our model above.

In a seminal paper, Hausman (1979) estimated a discrete choice model using 65 observations of consumer choices between air conditioner models, which vary in upfront cost and energy efficiency rating. Hausman framed his analysis as an estimate of an "implied discount rate" that rationalizes the demand system by assuming $\gamma = 1$. Hausman's paper, along with Dubin and McFadden's (1984) analysis of households' choices between heating systems, was the state of the art in this literature for 30 years. Both papers find real implied discount rates of 15 to 25 percent, which is higher than returns on stock market investments but not much different from real credit card interest rates, which were around 18 percent.

However, such analyses suffer from the problems introduced in the previous section. First, unobserved product attributes (which are analogous to ξ in our formal model) complicate the cross-sectional econometric approach. The coefficient on the present discounted value of energy costs is biased if energy efficient products have better or worse unobserved characteristics. For example, automobile prices actually *decrease* in fuel economy, as the more energy efficient vehicles are smaller and often have fewer luxury amenities. Furthermore, product prices will often be correlated with unobserved attributes, giving the usual simultaneity bias in estimating price elasticity. As in Berry, Levinsohn, and Pakes (1995), this issue can potentially be addressed by using instrumental variables, but the instruments available may be dissatisfying.

Working papers by Allcott and Wozny (2011), Busse, Knittel, and Zettelmeyer (2011), and Sallee, West, and Fan (2011) use an alternative approach to address the problem of unobserved attributes. These papers take a panel of used durable goods, condition on product fixed effects, and test how the relative prices of more energy efficient versus less efficient products change as energy price expectations vary over time. As an intuitive example of the identification strategy, notice that as expected gasoline prices rise, we should expect to see the market price of a three-year-old used Honda Civic increase relative to the price of a three-year-old Honda Accord, because the Civic is more energy efficient than the Accord. If market prices are not very responsive, this approach suggests that γ is small.

Relative to the other categories of evidence on the Energy Efficiency Gap, this approach is especially appealing because the fixed effects eliminate unobserved costs by construction. However, these analyses still suffer from the second problem, which is that they still require assumptions about the relevant discount rate, vehicle-miles traveled, and consumers' expectations of future gasoline prices (r, m, and p in our model above), and other factors. Allcott and Wozny's (2011) results tend to suggest that $\gamma < 1$, while Busse, Knittel, and Zettelmeyer's (2011) results tend not to support the hypothesis that $\gamma < 1$. The two analyses do agree that even if there is some investment inefficiency, the welfare losses would be relatively small. Allcott and Wozny show how to use discrete choice data to calculate the private welfare loss (equivalent to triangle *abf* in Figure 3 above). Their preferred estimate of γ suggests that investment inefficiencies in the auto market cause a welfare loss of about $1 billion per year and an increase in gasoline consumption of about 5 percent. Busse, Knittel, and Zettelmeyer's results tend not to suggest that there are investment inefficiencies, implying that there would be zero welfare losses. In either case, the welfare loss and gasoline consumption increase appear to be small relative to the total market size.

Investment Inefficiencies That Could Cause an Energy Efficiency Gap

In the previous section, we examined evidence on whether consumers and firms fail to exploit profitable energy efficiency investments. In this section, we reverse the perspective by specifying particular investment inefficiencies that might cause underinvestment in energy efficiency and assessing the empirical evidence on their magnitudes.

Imperfect Information

Imperfect information is perhaps the most important form of investment inefficiency that could cause an Energy Efficiency Gap. Two basic models of imperfect information are most relevant. In one model, consumers and firms may be unaware of potential investments in energy efficiency. For example, homeowners may not know how poorly insulated their home is and may not be aware of the opportunity to weatherize. Similarly, factory managers may not know about a new type of machine that could reduce their energy costs.

An alternative model resembles Akerlof's (1970) "lemons" model. Buyers know that different products, such as apartments, commercial buildings, or factory equipment, have different levels of energy efficiency, but these differences are costly to observe. Thus, they are not willing to pay more for goods that are in fact more energy efficient. For example, a renter evaluating a set of different apartments may be aware that there is a distribution of wall insulation quality and thus of resulting heating costs, but the renter will not be willing to pay more for a well-insulated apartment without taking the time to inspect the insulation.

There are three approaches to assessing the magnitude of imperfect information in the context of energy efficiency. The first approach is to test for market equilibria consistent with imperfect information. Several recent projects used this approach in the context of renter-occupied versus owner-occupied housing units. The theory is that because imperfectly informed renters will not be willing to pay more for energy efficient apartments, landlords have reduced incentive to invest in energy efficiency. Homeowners, on the other hand, do capture the benefits of improved energy efficiency, at least until they sell the property. Such a "landlord–tenant" agency problem implies that rental properties are less energy efficient than would be socially optimal. As an example of this approach, Davis (2010) studies the market penetration of refrigerators, dishwashers, light bulbs, room air conditioners, and clothes washers that have earned the U.S. government's "Energy Star" designation, meaning that they are relatively energy efficient. Conditional on observable characteristics, renters are 1 to 10 percentage points less likely to report having Energy Star appliances. The author calculates that if renters had the same energy efficient appliance ownership rates as owner-occupied homes, total energy bills in rental homes would be 0.5 percent lower.

The appliances that Davis (2010) considers make up only one-quarter of residential energy use. Heating and cooling represent close to one-half, meaning that insulation is perhaps a more important investment that could be subject to the landlord–tenant agency problem. Gillingham, Harding, and Rapson (2012) show that owner-occupied houses in California are 12 to 20 percent more likely to have insulation than rentals, conditional on other observable characteristics of the property, occupant, and neighborhood. Under the optimistic assumption that insulation reduces total energy demand by 10 percent, this implies that rental properties would use 1.2 to 2.0 percent less energy if insulated at the same level as owner-occupied properties. As both Davis (2010) and Gillingham, Harding, and Rapson (2012) note, conditional differences in appliance ownership between owners and renters are not ironclad evidence of a market failure, because preferences or housing stock could vary in unobservable ways.

If these estimates are assumed to be causal, how big is the investment inefficiency from the landlord–tenant agency problem? The magnitude is the number of affected households times the extent of the reduced energy efficiency. Of U.S. households, 29 percent are rental units where the renter pays energy bills (Murtishaw and Sathaye 2006). Multiplying this figure by several percent of total energy demand, to approximate the magnitude of the inefficiencies estimated above, implies that the landlord–tenant information problem might increase total residential energy use on the order of 1 percent. Thus, while the empirical evidence points to some inefficiency, it explains only a very small fraction of the purported Energy Efficiency Gap.[6]

An additional example of testing for imperfect information using equilibrium outcomes is to examine whether information disclosure increases the elasticity of energy-saving technical change with respect to energy prices. The idea is that consumers who are better informed about energy use will be more responsive to energy price changes when choosing between models of an energy-using durable. Therefore, firms with better-informed consumers will be more likely to offer more energy efficient models as energy prices rise. Newell, Jaffe, and Stavins (1999) show that the mean energy efficiency of room air conditioners and water heaters was

[6] A related agency problem is that some landlords pay the utilities, while tenants set the utilization levels for appliances and heating and cooling equipment. This problem is small, as only 4 percent of U.S. households are rentals with utilities included and demand for energy services is relatively inelastic. Levinson and Niemann (2004) estimate that energy costs are 1 to 1.7 percent higher when utilities are included in rent.

more responsive to energy prices after 1981 and 1977, respectively, the years when the federal government introduced energy efficiency labeling requirements for the two goods. While other factors might also have changed in these years, this finding suggests that the labeling requirements may have reduced the extent of imperfect information. However, this approach is not informative about the magnitude of any remaining investment inefficiency.

A second approach that can allow a direct assessment of the magnitude of imperfect information is to observe information sets through surveys. Turrentine and Kurani (2007) and Larrick and Soll (2008) use structured interviews and laboratory studies to show that consumers are not very good at calculating the gasoline costs for different automobiles. The 2010 Vehicle Ownership and Alternatives Survey (Allcott 2011a, 2012) adds nationally representative evidence on how accurately consumers perceive the financial value of energy efficiency. The data suggest that consumers are indeed imperfectly informed: over half of Americans mis-estimate the gasoline cost differences between the vehicle they own and their "second choice vehicle" by more than 40 percent. However, the errors run in both directions, and there is no clear evidence on whether the average consumer systematically underestimates or overestimates the energy cost savings from higher-fuel economy vehicles. Thus, while these analyses document imperfect information, none provides evidence that $\gamma < 1$.

A third approach to assessing the magnitude of imperfect information is to test for the effects of information disclosure on purchase decisions. This approach has the benefit of being based on observed choices in the marketplace, instead of beliefs stated on a survey. We are not aware of any large-scale randomized evaluations of energy efficiency information disclosure.

Inattention

Interventions that resemble information disclosure might change the buying patterns of consumers who are already well-informed. For example, Chetty, Looney, and Kroft (2009) find that despite the fact that consumers are well-informed about sales taxes, posting information about sales tax amounts in a supermarket changes buying patterns. This finding suggests the existence of another type of investment inefficiency, which behavioral economists call inattention.

The psychology of inattention starts by recognizing that choice problems have many different facets, and some of these facets are less salient at the time of choice even if they are potentially important to the utility that will later be experienced. When buying printers, for example, we might focus on the purchase price and fail to consider that replacement ink cartridges make up the bulk of the total cost. Inattentive consumers are misoptimizing: they fail to recognize opportunities to save money by choosing products with lower ancillary costs. Research in a variety of other non-energy settings is suggestive of inattention (Hossein and Morgan 2006; Barber, Odean, and Zheng 2005; Gabaix and Laibson 2006). It seems possible that some consumers might be inattentive to energy efficiency when purchasing energy-using durable goods.

Policy Implications

The available evidence on the size of an Energy Efficiency Gap is situation-specific, mixed, and often inconclusive. However, policymakers must make policy even in the absence of ironclad evidence.

The most important policy recommendation in this context, as in others, is to address market failures as directly as possible. In response to energy use externalities, a Pigouvian tax gives the first-best outcome. If agents are imperfectly informed and the government has an inexpensive information disclosure technology, an information disclosure approach should be used. Formulating policy becomes more challenging if the first-best solutions are not possible—when information disclosure is not fully effective, or when a Pigouvian tax is not politically feasible because of aversion to new taxes or to policies that explicitly regulate greenhouse gases. In this section, we examine the effects of energy efficiency policies, considered as a second-best alternative.

Energy Efficiency Subsidies and Standards as a Second-Best Approach to Pollution Abatement

Until now, we have set aside the uninternalized energy use externalities and focused on investment inefficiencies. We now examine the converse: say that no investment inefficiencies exist, but energy is priced below social cost (in our model, $\varphi > 0$). If Pigouvian taxes or cap-and-trade programs are politically infeasible, are energy efficiency subsidies and standards a relatively promising approach to pollution abatement?

When no investment inefficiencies exist, energy efficiency policies such as subsidies for energy efficient durable goods and minimum energy efficiency standards will have larger welfare costs per unit of pollution abated compared to the first-best Pigouvian tax for several reasons: First, subsidies and standards change relative prices for all consumers equally, while the Pigouvian tax provides a larger incentive for consumers with higher utilization to choose energy efficient capital stock. Second, the first-best policy must impose the right price on the utilization decision, which only the Pigouvian tax does. Third, it is difficult to calibrate the stringency of an energy efficiency standard or subsidy precisely, meaning that it will likely generate more or less carbon abatement than a Pigouvian tax set at the level of marginal damages. Energy efficiency policies in different sectors can also be miscalibrated against each other, causing inefficiency due to unequal marginal costs of abatement.

Of course, if these three theoretical factors were small in reality, then energy efficiency policies might be a reasonable second-best substitute for Pigouvian taxes. Several analyses have simulated the relative cost effectiveness of particular energy efficiency policies relative to Pigouvian taxes. Jacobsen (2010), for example, simulates automobile supply and demand and shows that Corporate Average Fuel Economy (CAFE) standards have a welfare cost of $222 per metric ton of carbon dioxide abated, compared to $92 per ton for a gas tax that generates the same amount of abatement. Krupnick, Parry, Walls, Knowles, and Hayes (2010) come to a similar qualitative conclusion. They compare the cap-and-trade provisions of the proposed Waxman–Markey climate change legislation to the legislation's energy efficiency provisions, which include standards for buildings, lighting, and appliances. The cap-and-trade, or an equivalent carbon tax, abates carbon dioxide at a welfare cost of $12 per ton. If there are no investment inefficiencies, the energy efficiency standards are five times more costly, or $60 per ton. This significantly exceeds the United States government's estimated social cost of carbon dioxide emissions, which is about $21 (Greenstone, Kopits, and Wolverton 2011).

These results forcefully argue that Pigouvian taxes or cap-and-trade programs are the most efficient way to address energy use externalities. Energy efficiency subsidies, CAFE standards, and other energy efficiency policies can also reduce energy use externalities, but in the absence of investment inefficiencies, they will often impose a significantly larger cost on the economy per unit of pollution reduction.

Energy Efficiency Subsidies and Standards as a Second-Best Approach to Correcting Investment Inefficiencies

The United States has long required energy-use information disclosure: for more than 30 years, retailers have been required to display fuel economy ratings for new vehicles and energy cost information for home appliances. However, consumers may not notice, understand, or pay attention to this information. If information disclosure or other direct solutions to market failures are not fully effective, how useful are energy efficiency subsidies and standards as a second-best approach to addressing investment inefficiencies?

Allcott, Mullainathan, and Taubinsky (2011) analyze this question when consumers are inattentive to energy costs. As we discussed earlier, their model shows that energy efficiency subsidies can increase welfare and, when consumers are sufficiently homogeneous, the first-best can be obtained. The intuition is straightforward: if consumers and firms underinvest in energy efficiency, subsidizing or mandating them to invest more can increase welfare. However, any corrective policies must be properly calibrated. For example, the vast majority of benefits in the U.S. government's cost–benefit analysis of Corporate Average Fuel Economy Standards derive from the assumption that the regulation corrects consumers' inattention to energy efficiency when buying autos.[7] However, Allcott and Wozny (2011), Fischer, Harrington, and Parry (2007), and Heutel (2011) use different models to show that the current and proposed CAFE standards are much more stringent than can be justified by even worst-case estimates of investment inefficiencies. Of course, if there are zero investment inefficiencies, then there are zero welfare benefits through this channel.

Heterogeneity in the investment inefficiency γ weakens the policy argument for subsidizing energy efficient goods, as Allcott, Mullainathan, and Taubinsky (2011)

[7] For more background, see the federal government's Regulatory Impact Analysis of the 2012–2016 CAFE standards (National Highway Traffic Safety Administration 2010) or the discussion in Allcott and Wozny (2011).

also show. In Figure 3 presented earlier, imagine that some consumers are on demand curve D' with $\gamma = 1$, while others are on demand curve D with $\gamma < 1$. Now, a subsidy moves the high-γ agents to point z, where they consume more energy efficiency than they would in the social optimum. This offsets the welfare gains from improving the decisions of the low-γ consumers. A key implication for policy analysis is that we must understand not just *how much* a policy increases sales of energy efficient goods, but *who* are the people induced to buy these goods. For example, even in a setting where the average consumer has $\gamma < 1$, energy efficiency subsidies might decrease total welfare if they are largely taken up by environmentalists and homeowners, who are more likely to be well-informed about energy efficiency and are not subject to a "landlord–tenant" agency problem.

This discussion highlights that energy efficiency policies are more likely to increase welfare if they target agents subject to the largest investment inefficiencies. Some existing policies do appear well-targeted. For example, households that use more energy than other comparable households are more likely to have low-cost energy conservation opportunities of which they are unaware, and many U.S. utilities now target energy conservation information to these relatively heavy users (Allcott 2011b; Ayres, Raseman, and Shih 2009). "Smart meters" that record hourly consumption, which as described in a companion paper by Paul Joskow in this symposium are increasingly being deployed across the United States, also provide information useful for targeting. For example, utilities can now identify households that use more energy on afternoon hours of particularly hot days, suggesting that they have energy inefficient air conditioners, and send them information on new energy efficient models.

Aside from heterogeneity in the investment inefficiency, consumers and firms also have substantial heterogeneity in other factors that affect demand for energy and for energy efficient capital stock. For example, the mild climate of Los Angeles compared to the more extreme weather of Chicago means that there is substantial variation in utilization of air conditioners and heating equipment, and residential retail electricity prices vary across the country from 4 to 30 cents per kilowatt-hour. As a result, national-level minimum efficiency standards for home appliances seem likely to decrease welfare for subsets of consumers with low prices and utilization and could increase welfare for high-price and/or -utilization consumers with investment inefficiencies. Ideally, standards could vary geographically to take account of this, targeting consumers that may have the most to gain. For example, building codes in states with extreme weather often require more insulation than building codes in mild climates. On the other hand, home appliance standards are set at the national level, and appliance manufacturers and retailers operate nationwide. The benefits of heterogeneous standards must be weighed against the costs of regulatory complexity.

Conclusion

Since the energy crises of the 1970s, many have made the "win-win" argument for energy efficiency policy: subsidies and standards can both address investment inefficiencies in the purchase of energy-using durable goods and reduce externalities from energy use. A reliance on observational studies of variable credibility and the possibility of unobserved costs and benefits of energy efficiency make it difficult to assess the magnitude of the Energy Efficiency Gap definitively. Nevertheless, the available evidence from empirical analyses of weatherization, demand-side management programs, automobile and appliance markets, the "landlord–tenant" agency problem, and information elicitation suggests that while investment inefficiencies do appear in various settings, the actual magnitude of the Energy Efficiency Gap is small relative to the assessments from engineering analyses.

Furthermore, it appears likely that there is substantial heterogeneity in investment inefficiencies across the population. Thus, targeted policies have the potential to generate larger welfare gains than general subsidies or mandates. Given this heterogeneity, policy analyses need to do more than assess how much a policy affects energy efficiency: they must also identify what types of consumers are induced to be more energy efficient.

We believe that this area is ripe for rigorous empirical research. Future research should utilize randomized controlled trials and quasi-experimental techniques to estimate the impacts of energy efficiency programs on heterogeneous consumer types and to address the challenges posed by unobserved costs and benefits. The economic insights from such research are potentially generalizable, and the policy implications are significant.

■ *We thank Michael Blasnik, Lucas Davis, Ken Gillingham, Matt Golden, Paul Joskow, Jim McMahon, Sendhil Mullainathan, Richard Newell, Jim Sallee, Alan Sanstad, Cass Sunstein, Rich Sweeney, and Catherine Wolfram for helpful discussions and feedback. We also thank the editors David Autor, Chad Jones, John List, and Timothy Taylor for helping to shape the article. Greenstone served as the Chief Economist at the Council of Economic Advisors when several energy efficiency policies were considered and enacted. Greenstone also holds a diversified portfolio of equities that include some energy providers. We gratefully acknowledge financial support from the Sloan Foundation, the MacArthur Foundation, and a research contract from the California Energy Commission to the Energy Institute at UC-Berkeley's Haas School for our research on the economics of energy efficiency.*

References

Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84(3): 488–500.

Allcott, Hunt. 2011a. "Consumers' Perceptions and Misperceptions of Energy Costs." *American Economic Review* 101(3): 98–104.

Allcott, Hunt. 2011b. "Social Norms and Energy Conservation." *Journal of Public Economics* 95(9–10): 1082–95.

Allcott, Hunt. 2012. "The Welfare Effects of Misperceived Product Costs: Data and Calibrations from the Automobile Market." Working Paper, New York University (January). Available at: https://files.nyu.edu/ha32/public/research.html.

Allcott, Hunt, Sendhil Mullainathan, and Dmitry

Taubinsky. 2011. "Externalities, Internalities, and the Targeting of Energy Policy." Working Paper, New York University (December). Available at: https://files.nyu.edu/ha32/public/research.html.

Allcott, Hunt, and Nathan Wozny. 2011. "Gasoline Prices, Fuel Economy, and the Energy Paradox." Working Paper, MIT (July). https://files.nyu.edu/ha32/public/research.html.

American Council for an Energy Efficient Economy (ACEEE). 1989. "The Potential for Electricity Conservation in New York State." Prepared for the New York State Energy RBD Authority, Research Report U891, September.

Anderson, Soren T., and Richard G. Newell. 2004. "Information Programs for Technology Adoption: The Case of Energy-Efficiency Audits." *Resource and Energy Economics* 26(1): 27–50.

Arimura, Toshi, Shanjun Li, Richard Newell, and Karen Palmer. 2011. "Cost-Effectiveness of Electricity Energy Efficiency Programs." *Resources for the Future Discussion Paper* 09-48-REV, April.

Auffhammer, Maximillian, Carl Blumstein, and Meredith Fowlie. 2008. "Demand-Side Management and Energy Efficiency Revisited." *Energy Journal* 29(3): 91–104.

Ayres, Ian, Sophie Raseman, and Alice Shih. 2009. "Evidence from Two Large Field Experiments that Peer Comparison Feedback Can Reduce Residential Energy Usage." NBER Working Paper 15386.

Barber, Brad, Terrance Odean, and Lu Zheng. 2005. "Out of Sight, Out of Mind: The Effects of Expenses on Mutual Fund Flows." *Journal of Business* 78(6): 2095–2120.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63(4): 841–90.

Blasnik, Michael. 2010. "Energy Models vs. Reality." Presentation. Southface Lecture Series. January 26.

Blumstein, Carl, and Steven E. Stoft. 1995. "Technical Efficiency, Production Functions and Conservation Supply Curves." *Energy Policy* 23(9): 765–68.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen. 2011. "Calorie Posting in Chain Restaurants." *American Economic Journal: Economic Policy* 3(1): 91–128.

Bovenberg, Lans, and Lawrence Goulder. 1996. "Optimal Environmental Taxation in the Presence of Other Taxes: General Equilibrium Analyses." *American Economic Review* 86(4): 985–1000.

Brown, Marilyn A., Mark D. Levine, Joseph P. Romm, Arthur H. Rosenfeld, and Jonathan G. Koomey. 1998. "Engineering–Economic Studies of Energy Technologies to Reduce Greenhouse Gas Emissions: Opportunities and Challenges." *Annual Review of Energy and the Environment* 23: 287–385.

Busse, Meghan, Christopher Knittel, and Florian Zettelmeyer. 2011. "Pain at the Pump: The Effect of Gasoline Prices on New and Used Automobile Markets." Working Paper, MIT, September. Available at Meghan Busse's website: http://www.kellogg.northwestern.edu/faculty/directory/busse_meghan.aspx#research.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99(4): 1145–77.

Davis, Lucas. 2008. "Durable Goods and Residential Demand for Energy and Water: Evidence from a Field Trial." *RAND Journal of Economics* 39(2): 530–46.

Davis, Lucas W. 2010. "Evaluating the Slow Adoption of Energy Efficient Investments: Are Renters Less Likely to Have Energy Efficient Appliances?" Energy Institute at Haas Working Paper 205, June.

Davis, Lucas, and Erich Muehlegger. 2010. "Do Americans Consume Too Little Natural Gas? An Empirical Test of Marginal Cost Pricing." *RAND Journal of Economics* 41(4): 791–810.

DeCanio, Stephen, and William Watkins. 1998. "Investments in Energy Efficiency: Do the Characteristics of Firms Matter?" *Review of Economics and Statistics* 80(1): 95–107.

Dixit, Avinash, and Robert Pindyck. 1994. *Investment under Uncertainty*. Princeton, NJ: Princeton University Press.

Dubin, Jeffrey, and Daniel McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(2): 345–62.

Dubin, Jeffrey, Allen Miedema, and Ram Chandran. 1986. "Price Effects of Energy-Efficient Technologies: A Study of Residential Demand for Heating and Cooling." *RAND Journal of Economics* 17(3): 310–25.

Duflo, Esther, Michael Kremer, and Jonathan Robinson. 2011. "Nudging Farmers to Use Fertilizer: Theory and Experimental Evidence from Kenya." *American Economic Review* 101(6): 2350–90.

Eto, Joseph, Suzie Kito, Leslie Shown, and Richard Sonnenblick. 1995. "Where Did the Money Go? The Cost and Performance of the Largest Commercial Sector DSM Programs." Lawrence Berkeley National Laboratory Working Paper No. LBL-38201.

Fischer, Carolyn, Winston Harrington, and Ian Parry. 2007. "Do Market Failures Justify Tightening Corporate Average Fuel Economy (CAFE) Standards?" *Energy Journal* 28(4): 1–30.

Foster, Andrew, and Mark Rosenzweig. 2010. "Microeconomics of Technology Adoption." Yale

University Economic Growth Center Discussion Paper 984.

Fowlie, Meredith, Michael Greenstone, and Catherine Wolfram. 2011. "An Experimental Evaluation of the Weatherization Assistance Program." Work in Progress.

Friedrich, Katherine, Maggie Eldridge, Dan York, Patti Witte, and Marty Kushler. 2009. "Saving Energy Cost-Effectively: A National Review of the Cost of Energy Saved through Utility-Sector Energy Efficiency Programs." ACEEE Report No. U092, September.

Gabaix, Xavier, and David Laibson. 2006. "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets." *Quarterly Journal of Economics* 121(2): 505–540.

Gillingham, Kenneth, Matthew Harding, and David Rapson. 2012. "Split Incentives and Household Energy Consumption." *Energy Journal* 33(2): 37–62.

Gillingham, Kenneth, Richard Newell, and Karen Palmer. 2006. "Energy Efficiency Policies: A Retrospective Examination." *Annual Review of Environment and Resources* 31: 161–92.

Gillingham, Kenneth, Richard Newell, and Karen Palmer. 2009. "Energy Efficiency Economics and Policy." *Annual Review of Resource Economics* 1: 597–620.

Greenstone, Michael, Elizabeth Kopits, and Maryann Wolverton. 2011. "Estimating the Social Cost of Carbon for Use in U.S. Federal Rulemakings: A Summary and Interpretation." MIT Department of Economics Working Paper No. 11-04, March.

Goldstein, David, Robert Mowris, Bart Davis, and Kari Dolan. 1990. *Initiating Least-Cost Planning in California: Preliminary Methodology and Analysis,* Natural Resources Defense Council and the Sierra Club, prepared for the California Energy Commission Docket No. 88-ER-8, Revised May 10.

Hassett, Kevin, and Gilbert Metcalf. 1993. "Energy Conservation Investment: Do Consumers Discount the Future Correctly?" *Energy Policy* 21(6): 710–16.

Hausman, Jerry. 1979. "Individual Discount Rates and the Purchase and Utilization of Energy-Using Durables." *Bell Journal of Economics* 10(1): 33–54.

Hausman, Jerry, and Paul Joskow. 1982. "Evaluating the Costs and Benefits of Appliance Efficiency Standards." *American Economic Review* 72(2): 220–25.

Heutel, Garth. 2011. "Optimal Policy Instruments for Externality-Producing Durable Goods under Time Inconsistency." NBER Working Paper 17083.

Hossain, Tanjim, and John Morgan. 2006. ". . . Plus Shipping and Handling: Revenue (Non) Equivalence in Field Experiments on eBay." *Advances in Economic Analysis and Policy* 6(2): Article 3.

Jacobsen, Mark. 2010. "Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity." September. http://econ.ucsd.edu/~m3jacobs/Jacobsen_CAFE.pdf.

Jaffe, Adam, and Robert Stavins. 1994. "The Energy Paradox and the Diffusion of Conservation Technology." *Resource and Energy Economics* 16(2): 91–122.

Joskow, Paul, and Donald Marron. 1992. "What Does a Negawatt Really Cost? Evidence from Utility Conservation Programs." *Energy Journal* 13(4): 41–74.

Koomey, Jonathan G., Celina Atkinson, Alan Meier, James E. McMahon, Stan Boghosian, Barbara Atkinson, Isaac Turiel, Mark D. Levine, Bruce Nordman, and Peter Chan. 1991. "The Potential for Electricity Efficiency Improvements in the U.S. Residential Sector." LBL-30477, Lawrence Berkeley Laboratory.

Krupnick, Alan, Ian Parry, Margaret Walls, Tony Knowles, and Kristin Hayes. 2010. *Toward a New National Energy Policy: Assessing the Options.* Resources for the Future Report, November. http://www.energypolicyoptions.org/wp-content/uploads/reports/RFF-Rpt-NEPI_Tech_Manual_Final.pdf.

Larrick, Richard, and Jack Soll. 2008. "The MPG Illusion." *Science* 320(5883): 1593–94.

Loughran, David, and Jonathan Kulick. 2004. "Demand-Side Management and Energy Efficiency in the United States." *Energy Journal* 25(1): 19–43.

Levinson, Arik, and Scott Niemann. 2004. "Energy Use by Apartment Tenants When Landlords Pay for Utilities." *Resource and Energy Economics* 26(1): 51–75.

McKinsey & Co. 2009. "Unlocking Energy Efficiency in the U.S. Economy." http://www.mckinsey.com/clientservice/electricpowernaturalgas/downloads/US_energy_efficiency_full_report.pdf.

Meier, Alan, and J. Whittier. 1983. "Consumer Discount Rates Implied by Purchases of Energy-Efficient Refrigerators." *Energy* 8(12): 957.

Meier Alan, Janice Wright, and Arthur H. Rosenfeld. 1983. *Supplying Energy through Greater Efficiency: The Potential for Conservation in California's Residential Sector.* Berkeley, CA: University of California Press.

Metcalf, Gilbert, and Kevin Hassett. 1999. "Measuring the Energy Savings from Home Improvement Investments: Evidence from Monthly Billing Data." *Review of Economics and Statistics* 81(3): 516–28.

Meyers, Stephen, James McMahon, and Michael McNeil. 2005. "Realized and Prospective Impacts of U.S. Energy Efficiency Standards for Residential Appliances: 2004 Update." Lawrence Berkeley National Laboratory Working Paper LBNL-56417.

Murtishaw, Scott, and Jayant Sathaye. 2006. "Quantifying the Effect of the Principal–Agent Problem on US Residential Energy Use." Lawrence Berkeley National Laboratory Working Paper Number 59773 Rev (August).

Nadel, Steven, and Howard Gelle. 1996. "Utility DSM: What Have We Learned? Where Are We Going?" *Energy Policy* 24(4): 289–302.

Nadel, Steven, and Kenneth Keating. 1991. "Engineering Estimates vs. Impact Evaluation Results: How Do They Compare and Why?" Research Report U915, American Council for an Energy-Efficient Economy. Available at: http://www.aceee.org/research-report/u915.

National Academy of Sciences. 1992. *Policy Implications of Greenhouse Warming: Mitigation, Adaptation, and the Science Base.* Washington, DC: National Academy Press.

National Highway Traffic Safety Administration (NHTSA), Office of Regulatory Analysis and Evaluation. 2010. *Corporate Average Fuel Economy for MY 2012–MY 2016 Passenger Cars and Light Trucks.* Final Regulatory Impact Analysis, March. National Highway Traffic Safety Administration, U.S. Department of Transportation.

Newell, Richard, Adam Jaffe, and Robert Stavins. 1999. "The Induced Innovation Hypothesis and Energy-Saving Technological Change." *Quarterly Journal of Economics* 114(3): 941–75.

Parry, Ian, David Evans, and Wallace Oates. 2010. "Are Energy Efficiency Standards Justified?" Resources for the Future Discussion Paper 10-59, November.

Rosenfeld, Arthur, Celina Atkinson, Jonathan Koomey, Alan Meier, Robert J. Mowris, and Lynn Price. 1993. "Conserved Energy Supply Curves for U.S. Buildings." *Contemporary Policy Issues* 11(1): 45–68.

Salle, James. 2010. "The Taxation of Fuel Economy." NBER Working Paper 16466.

Sallee, James. 2011. "Rational Inattention and Energy Efficiency." Work in progress, University of Chicago (June).

Sallee, James, Sarah West, and Wei Fan. 2011. "Do Consumers Recognize the Value of Fuel Economy? Evidence from Used Car Prices and Gasoline Price Fluctuations." Work in Progress, Macalester College (January).

Schweitzer, Martin. 2005. "Estimating the National Effects of the U.S. Department of Energy's Weatherization Assistance Program with State-Level Data: A Metaevaluation Using Studies from 1993 to 2005." Prepared for U.S. Department of Energy. Working Paper ORNL/CON-493, September.

Stoft, Steven. 1995. "The Economics of Conserved-Energy 'Supply' Curves." Program on Workable Energy Regulation (POWER) Working Paper 028, April.

Suri, Tavneet. 2011. "Selection and Comparative Advantage in Technology Adoption." *Econometrica* 79(1): 159–209.

Turrentine, Thomas, and Kenneth Kurani. 2007. "Car Buyers and Fuel Economy?" *Energy Policy* 35(2): 1213–23.

U.S. Bureau of Labor Statistics. 2007. "Table 1. Quintiles of Income before Taxes: Average Annual Expenditures and Characteristics, Consumer Expenditure Survey, 2007." ftp://ftp.bls.gov/pub/special.requests/ce/standard/2007/quintile.txt.

U.S. Department of Energy (DOE). 2011a. "Weatherization Assistance Program Goals and Metrics." Only available on the web at: http://www1.eere.energy.gov/wip/wap_goals.html.

U.S. Department of Energy (DOE). 2011b. "American Recovery & Reininvestment Act." Only available on the web at: http://www1.eere.energy.gov/recovery/.

U.S. Energy Information Administration (EIA). 2005. "Residential Energy Consumption Survey: Table US12: Total Consumption by Energy End Uses, 2005." http://www.eia.gov/emeu/recs/recs2005/c&e/summary/pdf/tableus12.pdf.

U.S. Energy Information Administration (EIA). 2010. *2009 Electric Power Annual.* DOE/EIA-0348(2009). Only available on the web at: http://www.eia.gov/cneaf/electricity/epa/epa_sum.html.

U.S. Energy Information Administration (EIA). 2011. Annual Energy Review "Table 2.1a: Energy Consumption Estimates by Sector, Selected Years, 1949–2010." http://www.eia.gov/totalenergy/data/annual/pdf/sec2_6.pdf.

U.S. Energy Information Administration (EIA). 2011. Residential Energy Consumption Survey. http://www.eia.gov/consumption/residential/.

U.S. Internal Revenue Service (IRS). 2011. "2009 Estimated Data Line Counts Individual Income Tax Returns." http://www.irs.gov/pub/irs-soi/09inlinecount.pdf.

Yergen, Daniel. 1979. *Energy Future: The Report of the Energy Project at the Harvard Business School.* New York: Random House.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 29–48

Creating a Smarter U.S. Electricity Grid

Paul L. Joskow

Starting in the late 1980s, federal and state governments began to restructure and deregulate some segments of the U.S. electric power industry. The basic idea was that the generation, transmission, physical distribution, and retail supply of electricity would be "unbundled" from one another. The physical distribution of electricity and the transmission of electricity would continue to be subject to state (distribution) and federal (transmission) regulation, while generation (wholesale competition) and retail supply (retail competition) would become competitive. To support this restructuring program, a number of regulatory and organizational changes were made or planned to create and manage wholesale power markets, transmission networks, and retail competition in an efficient manner.

These reforms spread quickly during the late 1990s. Then came the California Electricity Crisis (or the Western Electricity Crisis) of 2000–2001 (Joskow 2001; Borenstein 2002). The political reaction to this crisis put a virtual halt on additional states adopting restructuring and associated retail competition reforms. It also slowed efforts by the Federal Energy Regulatory Commission (FERC) to push forward its agenda to bring organized wholesale markets, integrating the efficient dispatch and pricing of generation supplied at different locations with the efficient allocation of scarce transmission capacity, to the entire country. FERC's efforts to rationalize the balkanized ownership and operation of transmission facilities by creating Regional Transmission Authorities (RTO) managed by Independent System Operators (ISOs) were also constrained. Today about one-third of the population has access to competitive retail supply alternatives, and

■ *Paul L. Joskow is President of the Alfred P. Sloan Foundation, New York City, New York. He is also the Elizabeth and James Killian Professor of Economics, Emeritus, Massachusetts Institute of Technology, Cambridge, Massachusetts. His e-mail address is ⟨joskow@sloan.org⟩.*

doi=10.1257/jep.26.1.29

about half of the generating capacity in the country is located in regions with organized competitive wholesale markets and transmission networks managed by independent system operators (Joskow 2006).

While efforts to refine the wholesale and retail competitive market reforms continue, public policy interest has now shifted to modernizing and expanding transmission and distribution networks. In particular, this paper focuses on efforts to build what policymakers call the "smart grid" by 1) stimulating investment to improve the remote monitoring and automatic and remote control of facilities on high-voltage transmission networks; 2) stimulating investment to improve the remote monitoring, two-way communications, and automatic and remote control of local distribution networks; and 3) installing "smart" metering and associated communications capabilities on customer premises so that customers can receive real-time price information and/or take advantage of opportunities to contract with their retail supplier to manage the consumer's demands remotely in response to wholesale prices and network congestion. While the smart grid is the focus of this paper, there are other important areas for modernizing and expanding transmission networks, including stimulating investment in new transmission capacity, especially "long distance" transmission facilities that span multiple states, and better integrating electricity demand into wholesale power markets.

A recent Electric Power Research Institute (2011a, p. 1-1) report uses the following definition of the smart grid:

> The term "Smart Grid" refers to the modernization of the electricity delivery system so that it monitors, protects, and automatically optimizes the operation of its interconnected elements—from the central and distributed generator through the high-voltage transmission network and the distribution system, to industrial users and building automation systems, to energy storage installations, and to end-use consumers, and their thermostats, electric vehicles, appliances, and other household devices.

Current "smart grid" initiatives are driven by a number of potential benefits. The EPRI (2011a, p. 1-1) report correctly notes: "The present electric power delivery infrastructure was not designed to meet the needs of a restructured electricity marketplace, . . . or the increased use of renewable power production." The reference to a "restructured marketplace" emphasizes that a smarter grid can facilitate wholesale and retail competition in the supply of power, as well as the need to accelerate replacement of an aging transmission and distribution infrastructure and to conserve on meter reading and other network operating costs. The reference to renewable power points out that a smart grid may be needed if solar, wind, geothermal, and other renewable energy technologies are to make a sizable contribution to national electricity needs as well as engage with demand-side issues like charging electric vehicle batteries or encouraging consumers to use electricity more efficiently.

The American Recovery and Reinvestment Act of 2009 (ARRA) provided $4.5 billion in funds for smart grid demonstration and technology deployment projects, including various analyses of consumer behavior in response to the installation of "smart meters," discussed at ⟨http://www.smartgrid.gov/federal_initiatives⟩, a website sponsored by U.S. Department of Energy. About 140 projects have been funded under these programs with about $5.5 billon of matching funds from utilities and their customers. Several states have adopted regulations that require utilities to install smart meters and make other smart grid investments, while others have started more modestly with pilot programs. The costs of these efforts are typically recovered through regulated prices for physical distribution services. The federal funds have certainly accelerated activity on smart grid projects around the country, and these financial incentives have been reinforced by state mandates and pilot programs. Since it is unlikely that federal subsidies for smart grid investments will be sustained at their recent ARRA level, the performance of the projects supported with these funds and the experience with state mandates and pilot programs will be a powerful influence on whether and how fast smart grid investments continue in the future.

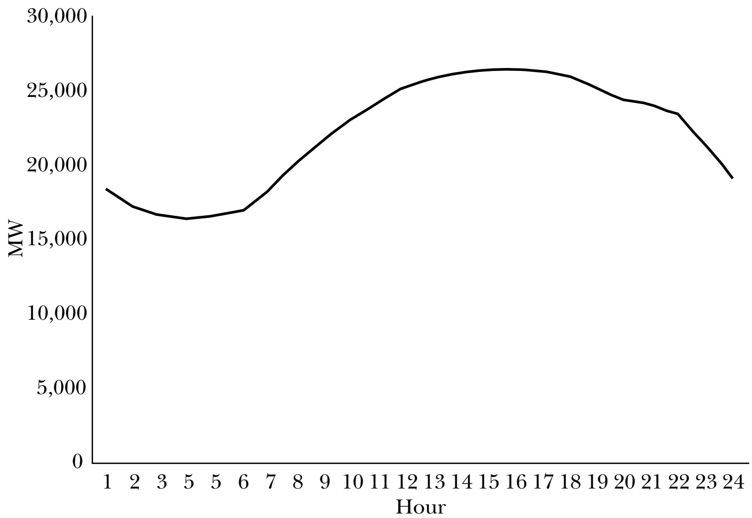
In what follows, I will examine the opportunities, challenges, and uncertainties associated with investments in "smart grid" technologies at each of the traditional components of the grid. I start by discussing some basic electricity supply and demand, pricing, and physical network attributes that are critical for understanding the opportunities and challenges associated with expanding deployment of smart grid technologies. I then discuss issues associated with the deployment of these technologies at the high voltage transmission, local distribution, and end-use metering levels. I will not discuss "behind the meter" technologies that may be installed inside of homes and businesses in response to the availability of smart grid capabilities, smart metering, and variable pricing.

Attributes of Electricity Markets

The demand for electricity varies widely from hour to hour, day to day, and month to month. Electricity demand is typically highest during the daytime hours and lowest at night. It tends to be very high on unusually hot or unusually cold days and is lowest at night on mild spring and fall days. Demand typically reaches its highest levels during only a few hours each year. There is also a minimum "base" aggregate demand that is sustained through the entire year. Figure 1 displays the levels of demand or "load" at different times of the day in New England on July 7, 2010. The peak demand is 60 percent higher than the lowest demand on that day. Figure 2 depicts the associated spot prices for electricity at each hour on that day, which I will discuss presently.

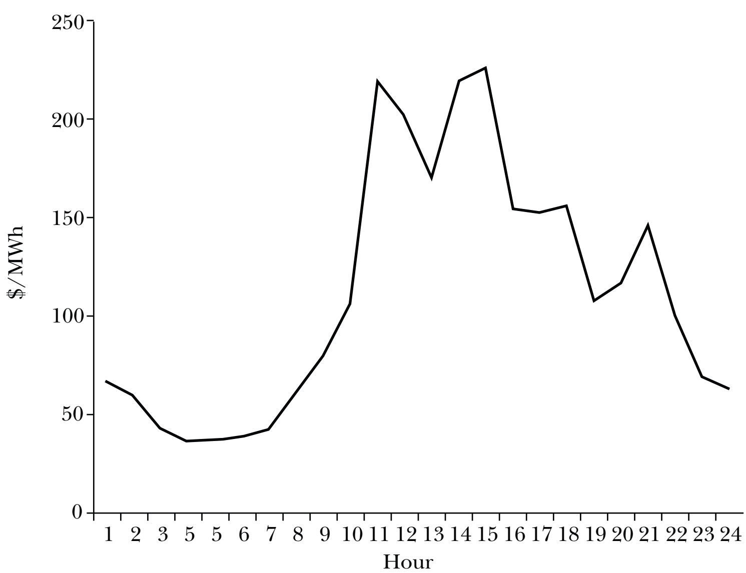
Electricity cannot be stored economically for most uses with current technologies (except in special applications where batteries, pumped storage, compressed air, and the like may be economically attractive). In electricity markets, physical

Figure 1
Real-Time Demand for Electricity, July 7, 2010
(in megawatts)



Source: Constructed from data from the New England ISO at ⟨http://www.iso-ne.com⟩.

Figure 2
Real-Time Energy Prices, July 7, 2010
(dollars per megawatt hour)



Source: Constructed from data from the New England ISO at ⟨http://www.ne-iso.com⟩.

inventories are not generally available to balance supply and demand in real time, and "stockouts" mean rolling blackouts or a larger uncontrolled system collapse (Joskow and Tirole 2007). On an electricity grid, supply and demand must be balanced continuously to maintain a variety of physical network criteria—like frequency, voltage, and capacity constraints—within narrow bounds. Electricity is the ultimate "just in time" manufacturing process, where supply must be produced to meet demand in real time.

These considerations have implications for the variations in the spot price of electricity in an unregulated wholesale electricity market and the shadow price of electricity in a traditional regulated environment that relies on an economic dispatch curve based on estimates of marginal generating costs. As demand increases, "dispatchable" generating capacity—first "base load," then "intermediate," then "peaking" capacity—with higher and higher marginal operating costs, is called to balance supply and demand (Turvey 1968; Boiteux 1964a; Joskow and Tirole 2007). As a result, wholesale market prices that reflect short-run marginal costs of generation are generally high when demand is high and low when demand is low, reflecting the marginal cost of the generation supplies needed to meet demand at different points in time. During unusually high-demand periods, supply and demand may (theoretically) be rationed on the demand side. When unexpected outages occur due to generation supply constraints or network failures, electricity consumers bear costs typically measured as the Value of Lost Load or VOLL (Stoft 2002; Joskow and Tirole 2007).

As noted, Figure 2 displays the variations in wholesale spot prices in New England associated with the variations in demand displayed in Figure 1 for a hot day in July 2010. The highest price is five times the lowest price on that day. More extreme price variability has been observed under more extreme weather conditions, though there is a $1,000 per megawatt hour cap placed on spot prices for energy in most areas ($3,000 in Texas).[1]

The prices in Figure 2 are *wholesale* spot prices. However, most *retail* residential and small commercial consumers are charged a retail price per kilowatt hour that does not vary dynamically with the time they consume electricity. As a result it does not reflect the wide variations in wholesale prices and the marginal cost of generating electricity. This is the case because traditional residential and small commercial users of electricity have meters that record only aggregate consumption between monthly or semi-monthly readings.[2] In some states, residential and small commercial consumers can opt for time-of-use meters, which charge different pre-set prices during predetermined "peak" and "off-peak" periods (for example, daytime and nighttime prices), based on averages of historical prices. While these

[1] The price caps are generally thought to be well below the Value of Lost Load in most circumstances and this raises other issues for efficient short-run and long-run performance of competitive wholesale markets (Joskow 2005).

[2] In a few cases, the largest retail consumers were billed based on prices that did vary more or less with variations in wholesale market prices (Mitchell, Manning, and Acton 1978, pp. 9–16).

time-of-use retail prices somewhat more accurately reflect variations in marginal generation costs and wholesale market prices, the relationship between actual retail prices and actual wholesale prices is necessarily very rough indeed, and penetration of time-use retail pricing has been low.

Enhancing High Voltage Transmission Systems

High voltage transmission networks are central to the operation of a modern electric power system: they make it possible to meet locationally dispersed demand with locationally dispersed generation in an efficient and reliable manner. High voltage alternating current (AC) networks are not switched networks—like a traditional railroad or telephone network—that is, power generated at point A does not flow to a specific customer located at point B. Electricity flows on an AC power network according to physical laws known as Kirchoff's laws and Ohm's law (Clayton 2001; Stoft 2002; Hogan 1992; Joskow and Tirole 2000). To drastically oversimplify, electricity produced on an AC electric power network distributes itself to follow the paths of least resistance. Transmission networks can also become congested in multiple locations, which may not lie on the "path" between buyer B and seller A, as suppliers of relatively low-cost generation seek to use the network to sell power to areas with higher cost generation. Network congestion is reflected in differences in wholesale market prices for electricity (or in shadow prices where wholesale markets with locational pricing have not been created) at different locations on the network (Hogan 1992; Joskow and Tirole 2000).

Each of the three high voltage AC networks covering the continental United States experiences significant congestion during certain hours of the year, including many "off-peak" hours, although as far I know the costs of congestion have never been quantified systematically for the entire country.[3] A natural approach to measuring the magnitude and costs of congestion is to use the differences in locational wholesale prices over time. For example, Table 1 displays the average spot wholesale prices during peak hours at different locations on the Eastern Interconnection on that same day in July 2010. It should be clear that on July 7, 2010, power was not flowing from one location to another on the Eastern Interconnection to arbitrage away large differences in wholesale spot prices; the ability of the network to transfer power from one location to another was constrained by scarce transmission capacity.

This congestion and lack of wholesale locational price arbitrage arises for three primary reasons: First, the transactions costs for moving power from the North,

[3] These costs, or at least the congestion rents, are quantified for the Regional Transmission Organizations (RTO) and Independent System Operators that have markets based on a locational marginal price market design. For example, in the PJM RTO, region, the independent market monitor estimated congestion costs to be as high as $2 billion in 2005, with substantial year-to-year variation, in his State of the Market Report for 2006 (PJM 2007, p. 39).

Table 1
Day-Ahead Peak Period Prices for Delivery July 7, 2010

Location	*$/MWh*
Boston (Massachusetts Hub)	117.75
New York City (Zone J)	138.50
Buffalo (Zone A)	79.00
Virginia (Dominion Hub)	107.75
Chicago (Illinois Hub)	68.75
Minneapolis (Minnesota Hub)	42.50
Florida	37.00

Source: Megawatt Daily, July 7, 2010, p. 2.

West, and South to New York City are high, requiring transactions with multiple Regional Transmission Organizations, Independent System Operators, and other balancing authorities with different market designs, settlement rules, and transmission service prices. Second, system operators place a very high value on reliability, which means maintaining "contingency" margins to be prepared for unanticipated events in neighboring areas which might affect their area. Third, most system operators have inadequate monitoring, communication, and control equipment on their high voltage network—an inadequate ability to "see" the state of neighboring networks—so they enforce higher contingency margins than would be necessary if they had better information and a wider span of control.

The benefit of these high contingency margins is that the U.S. electric transmission system is presently very reliable. While good comprehensive numbers are not available, it is extremely rare that retail consumers lose power because of failures of equipment or operating errors on the high voltage transmission system. EPRI (2011a, p. 2.1) estimates that U.S. power systems achieve 99.999 percent reliability at the high voltage (bulk) transmission network level and that over 90 percent of the outages experienced by retail customers are due to failures on the distribution system, not the transmission system (p. 6.1). However, when a rare major failure does occur on the high voltage transmission network, as with the 2003 Northeast Blackout when 50 million customers suffered power outages that lasted up to a couple of days, the associated costs can be high. (The 2003 Northeast Blackout was due in part to poor communications between system operators of interconnected control areas.)

It is widely accepted that there has been underinvestment in monitoring, communications, and control equipment on the high voltage transmission network to meet the needs of supporting efficient wholesale power markets, squeezing more effective capacity from existing transmission facilities, and achieving policy goals for renewable energy from grid-based wind and solar generating systems (for discussion, see EPRI, 2011a, chap. 5; New York ISO, ⟨http://www.nyiso.com/public/energy_future/issues_trends/smart_grid/index.jsp⟩; U.S. Department of Energy ⟨http://www.oe.energy.gov/⟩). EPRI (2011a, p. 5.1) recognizes that while it is hard to estimate with precison the costs of upgrading the high voltage transmission system with this "smart" equipment, EPRI estimates that the total investment cost is $56–$64 billion. EPRI also concludes the investments in improved monitoring of high voltage transmission networks represent the most cost-effective category of smart grid investments. Investments in this category also represent about 20 percent of the total cost of EPRI's defined Smart Grid program. This is qualitatively consistent with my own assessment.

These smart grid investments at the high voltage transmission level are likely to have even higher returns as "intermittent" generating capacity like high voltage grid-connected wind and solar generating capacity grows (local photovoltaic facilities on the roofs of homes create related challenges for distribution networks—see below). High voltage grid-based wind and solar installations supply electricity intermittently. This means that their output is driven by wind speed, wind direction, cloud cover, haze, and other weather characteristics rather than by supply and demand conditions and wholesale market prices. As a result, their output typically cannot be controlled or economically dispatched by system operators based on economic criteria in the same way as traditional electricity generation technologies (Joskow 2011a, b). Since wind and sun intensity vary widely and quickly, output of intermittent generating units can vary widely from day to day, hour to hour, minute to minute, and location to location. To balance supply and demand continuously when there is significant intermittent generation on the high voltage network requires that system operators have the capability to respond very quickly to rapid changes in power flows at different locations on the network by holding more dispatchable generation in operating reserve status and having the capability to monitor and adjust the configuration of power flows on the transmission network to balance supply and demand continuously while minimizing costs.[4]

Smart grid investment on the high voltage network has only a limited ability to increase the effective capacity of transmission networks. A large increase in transmission capacity, especially if it involves accessing generating capacity at new locations remote from load centers, requires building new physical transmission capacity. However, building major new transmission lines is extremely difficult. The U.S. transmission system was not built to facilitate large movements between interconnected control areas or over long distances; rather, it was built to balance supply and demand reliably within individual utility (or holding company) service areas. While the capacity of interconnections have expanded over time, the bulk of the price differences in Table 1 are due to the fact that there is insufficient transmission capacity to move large amounts of power from, for example, Chicago to New York City. The regulatory process that determines how high voltage transmission capacity (and smart grid investments in the transmission network) is sited and paid

[4] These network issues associated with intermittent generating capacity are different from issues related to the proper comparative valuation of intermittent and dispatchable generating technologies (Joskow 2011a, b); Borenstein (2008) applies methods compatible with those in Joskow (2011a, b) to derive the cost per kilowatt hour supplied and the cost per ton of CO_2 displaced by substituting solar for fossil-fuel generation expected to result from California's rooftop solar energy subsidy program.

for in regulated transmission prices is of byzantine complexity (Joskow 2005). It is clear, however, that the combination of FERC cost allocation policies, the requirement to receive siting permits from each state in which a new transmission line is located, and not-in-my-backyard political constraints hinder efficient investment in long distance transmission lines. FERC has been trying to resolve the issue of "who pays" and "how much" for new transmission lines for years, most recently promulgating Order 1000 in July 2011. This rule has many constructive features, but it will take several years to see how and to what extent it is implemented. Nor does that rule address state siting requirements or NIMBY constraints. Congress gave the Department of Energy authority to designate National Interest Electric Transmission Corridors to respond to the diffusion of siting authority among many states, but the DOE's procedures have been rejected by the courts (Watkiss 2011). The best solution to the siting problems would be to move regional transmission planning authority from the states to FERC. However, the political barriers to such a change are enormous. Thus, underinvestment in multistate high voltage transmission facilities is likely to continue to be a problem for many years.

Automating Local Distribution Networks

The smart grid technologies being deployed on local distribution systems include enhanced remote monitoring and data acquisition of feeder loads, voltage, and disturbances; automatic switches and breakers; enhanced communications with "smart" distribution substations, transformers, and protective devices; and supporting communications infrastructure and information processing systems. Smart grid investments in local distribution networks offer a variety of potential gains: to reduce operation and maintenance costs (goodbye meter readers, manual disconnects, and responses to nonexistent network outages); to improve reliability and responses to outages; to improve power quality (for example, to eliminate very short disruptions in voltage or frequency); to integrate distributed renewable energy sources, especially solar photovoltaic systems installed at customer locations that produce power intermittently and can lead to rapid and wide variations in the (net) demand placed on the distribution network; to accommodate demands for recharging of the electric vehicle of the future; to deploy "smart meters" that can measure customers' real-time consumption and allow for dynamic pricing that reflects wholesale prices; and to expand the range of products that competing retail suppliers of electricity can offer to customers in those states that have adopted retail competition models.

The U.S. Department of Energy has supported about 70 smart grid projects involving local distribution systems on a roughly 50/50 cost sharing basis, with details available at ⟨http://www.smartgrid.gov/recovery_act/tracking_deployment/distribution⟩. However, a full transformation of local distribution systems will take many years and a lot of capital investment. Are the benefits likely to exceed the costs?

In the only comprehensive and publicly available effort at cost–benefit analysis in this area, the Electric Power Research Institute (2011a) estimates that deployment (to about 55 percent of distribution feeders) would cost between $120–$170 billion, and claims that the benefits in terms of greater reliability of the electricity supply would be about $600 billion (both in net present value).[5] Unfortunately, I found the benefit analyses to be speculative and impossible to reproduce given the information made available in EPRI's report.

According to EPRI (2011a, page 6.1), over 90 percent of the electricity supply outages experienced by retail electricity consumers occur because of failures on the local distribution network rather than the transmission network. These failures may be caused by wind and storms, tree limbs falling on overhead distribution lines, icing up of distribution equipment, overloads of the local distribution network, failures of low-voltage transformers and breakers due to age or poor maintenance, cars that crash into poles and knock down distribution equipment, flooding of underground distribution, excessive heat, natural aging, and so on. No matter how smart we make local distribution systems, power outages will arise from many of the natural causes on this list, especially in areas that rely on overhead (rather than underground) distribution lines. Using standard measurement criteria from the IEEE (Institute of Electrical and Electronics Engineers), which exclude certain planned and weather-related outages, the average residential household has (rounding to simplify the calculation) 1.5 unplanned outages per year with an average outage duration of about 100 minutes per year (Power Engineering Society, 2006). Accordingly, the average residential customer experiences about 150 minutes of unexpected outages per year or 10.5 percent of one day per year. When I compare EPRI's estimates of the benefits from greater reliability with typical estimates of Value of Lost Load—for example, $5,000 to $10,000 per megawatt hour lost—the EPRI estimates of reliability benefits appear much too high.

Very short voltage drops and electrical transients that appear almost as flickers of lights (poor "power quality")—potentially create significant problems for very sensitive digital equipment. Investments in smart distribution grid technology can reduce these transients, but at significant cost. The value of reducing these transients also varies widely among customers. Having to reset one's clock is less costly than maintaining backup facilities for a critical server or data storage system in the event of disruption or damage from a voltage spike, as a financial management firm might have to do. In crafting a response, we must address the question of whether investments to improve power quality should be made for everyone, or whether they should be made "behind the meter" by those who value power quality highly? This issue would benefit from more independent empirical evidence and analysis.

[5] More specifically, nearly half of the overall benefits ($445 billion in net present value) for EPRI's (2011a) entire smart grid program are attributed to "reliability," which appears to be shorthand for reliability and power quality. There is another benefit category called "security" (benefits of $151 billion in net present value), which seems to be a subset of "reliability." Adding these gives the total of roughly $600 billion in the text.

Of more pressing concern are the new demands that may be placed on at least some distribution systems by distributed generation, primarily solar photovoltaic systems, and by the potential future need to recharge plug-in electric vehicle batteries. Several states are promoting solar photovoltaic technology with large subsidies (Borenstein 2008). Due to the wide variability in the output of photovoltaic technologies, sometimes over short time periods (NERC 2009, pp. 27–29), and related variability of net demand (which could be net *supply* if there are enough photovoltaic facilities and it is very sunny) that consumers place on the distribution system, photovoltaic technologies will place new stresses on local distribution feeders where they are installed. Better remote real-time monitoring and remote and automatic control capabilities, data acquisition and analysis of the state of the distribution system, and automatic breakers and switches will be required to accommodate significant quantities of these resources efficiently and safely.

The rate of new photovoltaic installations will vary widely from distribution feeder to distribution feeder and from state to state because of differences in subsidy policies and the relative economic attractiveness of photovoltaic investments. This variation suggests a targeted approach to local distribution system automation: focus first on areas where distributed generation, and the related stress on specific distribution feeders, will happen sooner.

The potential future demands placed on the local electric distribution system by plug-in electric vehicles raise similar issues. In 2010, out of 11.6 million total car sales, there were at most 3,000 pure electric plug-in electric vehicles sold and about 275,000 plug-in hybrids. The future path of electric vehicle sales depends on many factors: the price of gasoline, subsidies for electric vehicles, technological change affecting battery life and costs, new emissions standards, reductions in electric vehicle costs, and consumer behavior. Forecasts of the fraction of new vehicles that will be electric plug-ins by 2035 varies from less than 10 percent to over 80 percent. The U.S. Energy Information Administration (EIA 2011a, p.72; see also EPRI 2011b, chap. 4) forecasts a market share of light-duty vehicles of only 5 percent for plug-in and all-electric vehicles in 2035 in its reference case. The National Research Council (Committee on Assessment of Resource Needs for Fuel Cell and Hydrogen Technologies, 2010, p. 2) concludes that a realistic estimate is that by 2030 about 4.5 percent of the national light duty vehicle fleet will be plug-in electrics—with a maximum possibility of about 13 percent.

Along with the number of plug-in cars, the load placed on the distribution system will depend on the attributes of the car batteries and charge-up time selected by vehicle owners. Shorter charging times at higher voltages can place very significant loads on local distribution networks even with modest electric vehicle penetration (Browermaster 2011). An interesting possibility arises here. The demand on portions of the *local distribution system* in areas where electric vehicle sales may be concentrated—say, in places like Berkeley, California or Cambridge, Massachusetts—could peak at night, even when prices in the much broader wholesale power market are low. The possibility that the marginal cost of electricity distribution service on some distribution feeders could peak at times when aggregate system demand for power and associated power prices in the wholesale market are low suggests that more thought should be given to altering the pricing of electricity distribution service. Distribution service prices are now based on a flat per kilowatt hour rate and do not vary with the marginal costs of distribution service, which is driven largely by the peak demand on distribution feeders. Distribution service prices would more closely reflect the associated marginal cost of this service if we moved, instead, to a distribution charge based at least partially on individual customers' peak load on the local distribution system.

These considerations all lead to the conclusion that phasing in the installation of smart grid technology, targeting investment where it is likely to be most needed quickly, and collecting data using a controlled experimental framework to evaluate its costs and benefits would be a sensible strategy.

I offer one caveat to this conclusion. Many U.S. distribution systems are aging, and utilities are embarking on large distribution network replacement programs. These investments are long-lived, and it makes sense for these programs to take advantage of the most economical modern distribution technologies. In many cases, forward-looking investment optimally should deploy much more automation and communication technologies than immediately needed, even if deployment of distributed generation and electric vehicles is expected to be slow.

Smart Meters and Dynamic Pricing Incentives

It is not unusual for the incremental generating capacity held by a utility or RTO to meet the peak demand on its system during the 100 highest demand hours each year (1.1 percent of the total hours) to account for 10 to 15 percent of the generating capacity on its system. This is the direct consequence of the wide variability of demand (especially in response to extreme weather conditions), the failure of retail prices to reflect the true marginal cost of supply under these extreme conditions, and the utility practice of building enough generating capacity to meet demand even under very extreme low-probability demand states. Accordingly, using appropriate prices to provide consumers with an incentive to cut peak demand during a small number of hours can reduce generating costs significantly in the long run. Retail prices that are not tied to variations in wholesale prices inefficiently increase the level of peak demand by underpricing it, and may also discourage increased demand during off-peak hours by overpricing it.

The idea of moving from time-invariant electricity prices to "peak-load" pricing where prices are more closely tied to variations in marginal cost has been around for at least 50 years (Boiteux 1964b, c; Turvey 1968; Steiner 1957; Kahn 1970, pp. 63–123). The application of the theory in practice has lagged far behind, especially in the United States. (Mitchel, Manning, and Acton (1978) discuss early developments in other countries.) There is evidence from the well-designed experiments with time-of-use pricing of electricity in the 1970s showing consumers do respond more or less as expected to price incentives (Aigner 1985), in the sense that higher prices lead to less consumption. The magnitude of the estimated responses varies widely though and does not reflect potential adaptations in the attributes of the appliance stock resulting from widespread deployment. Moreover, a 2010 survey indicated that only about 1 percent of residential customers were on time-of-use rates (FERC 2011, pp. 28, 99).

"Smart meters" can record *real-time* consumption of electricity. They also can have two-way communications capabilities allowing for real-time retail prices tied to variations in wholesale prices (and a number of variations on this theme), and could lead to remote control of customer demand by allowing the retail supplier or the customer to adjust appliance utilization inside the customer's home or business. For example, a customer might program the air conditioning (or have the utility program it) to turn off, at least intermittently, when electricity prices reach a certain level. Real-time pricing to reflect variations in wholesale market prices can increase efficiency, at least when it is applied to larger customers (Borenstein 2005). In addition, real-time pricing can stimulate innovations in appliances and equipment inside the home or in business to make them capable of responding more easily to changes in real-time prices and load management arrangements made with retail suppliers.

The traditional arguments for not introducing real-time pricing were: 1) the meters and billing costs would be so costly that residential and small commercial customers would not benefit from them; 2) retail consumers would not understand or effectively utilize complex rate designs; and 3) changing rate designs would lead to large redistributions of income reflecting the wide variations in consumption patterns across individuals and decades-old mechanisms for allocating costs among types of customers and within customer classes (Borenstein 2007a, b).

At least some of these arguments are increasingly being questioned, and solutions being contemplated. Metering and communications technology have moved forward with more capabilities and lower costs. Today's more advanced smart meters use two-way communication and capabilities for active real-time interaction between the distribution system and the customer: they can record consumption at least once each hour, can be turned on and off remotely, can support the introduction of dynamic retail prices that are closely tied to dynamic wholesale market prices, and can control the utilization of appliances remotely in a way that facilitates active demand-side management of the electrical grid. In addition, the information available through smart meters can inform the distribution company about variations in demand and equipment outages on the distribution grid instantly, thus creating synergies between "smart meters" and smart distribution grid investments. Variations on full real-time pricing, in particular "critical peak pricing," are easier for consumers to understand and provide much better incentives than flat rates. Nevertheless, relatively few advanced "smart meters" had been installed and used effectively in the United States, although the number is now increasing at a rapid rate (U.S. Energy Information Administration 2011b) as a result of federal subsidies and state mandates. As many as 8.7 million smart meters have now been installed at residential and small commercial locations, about 6 percent of the total—though the definition of what counts as a smart meter varies and experience with them is limited (St. John 2009).

Analysis of the costs and benefits of large-scale deployment of smart meters must look at both changes in consumer surplus and changes in the cost of supplying electricity and metering its consumption. On the demand side, one needs to be able to measure the demand elasticities and cross-elasticities for a very diverse population of consumers who have different appliance stocks, live in homes of widely varying sizes, experience different weather conditions, face different levels and structures of incumbent electricity tariffs, have different incomes, and consume different quantities of electricity each month. An added complexity is that it would be implausible to measure *long-run* demand elasticities taking the current attributes of appliances and equipment as given, because the appliance stock and opportunities to adjust energy use will change to take advantage of the new incentives if smart meters and dynamic pricing are widely used.

On the supply side, there are questions about how much all of this whizzy smart grid technology will cost and, as always, there is the need to measure the effects on generation, distribution, and transmission costs. Measuring incremental metering costs is not easy. The many different vendors of smart meters sell meters with different functionalities and different communications methods. Moreover, buying and installing the meters is only part of the cost: communications systems must be built to integrate smart meter information with automated distribution network capabilities; a new information technology infrastructure for data acquisition, analysis, and billing created and installed; customer service personnel retrained to respond to questions about more complex rate structures; and investments made in complementary distribution system upgrades. On the other hand, smart meters should also save operating costs, primarily by reducing meter-reading costs (especially for systems that have not already installed the first-generation one-way communication meters). We also know that as a theoretical matter, setting retail prices to reflect marginal supply costs will increase overall efficiency with which electricity is consumed and supplied. But is this efficiency gain large enough to cover the additional costs of smart meters and associated information and automated distribution technology, both in the aggregate and for customers with different utilization characteristics? I do not think that this question has yet been answered satisfactorily or the public adequately convinced that the answer is likely to be affirmative.

A large number of U.S. utilities began offering time-of-use and interruptible pricing options for large commercial and industrial customers during the 1980s, either as a pilot program or as an option (for example, Barbose, Goldman, Bharvirkar, Hopper, Ting, and Neenan 2005). More recently, a number of states have introduced pilot programs for residential (household) consumers that install smart meters of various kinds, charge prices that vary with wholesale prices, and observe demand. For example, Taylor, Schwarz, and Cochell (2005) estimate hourly own- and cross-price elasticities for industrial customers on Duke Power's optional real-time rates and find large net benefits for these customers. Faruqui and Sergici (2010) summarize the results of 15 earlier studies of various forms of dynamic pricing, including time-of-use pricing, peak pricing, and real-time pricing. Faruqui and Sergici (2011) analyze the results of a dynamic pricing study performed by Baltimore Gas & Electric using treatment and control groups drawn from a representative group of households. Wolak (2006) analyzes a peak pricing experiment in Anaheim, California. Wolak (2010) analyzes a pilot program using peak pricing in Washington, D.C. Allcott (2011) analyzes data from the Chicago Energy Smart Pricing Plan that began operating in 2003. Faruqui (2011) summarizes the reduction in peak load from 109 dynamic pricing studies, including those that use time-of-use pricing, peak pricing, and full real-time pricing, and finds that higher peak period prices always lead to a reduction in peak demand. However, the reported price responses across these studies vary by an order of magnitude, and the factors that lead to the variability of responses have been subject to very limited analysis.

Before discussing what conclusions can be drawn from this evidence, a few warnings seem appropriate. First, there is wide variation in the design of the pilot/experimental studies and the variation in prices included in them. Just looking at the magnitude of the response without more information is not adequately informative. Second, essentially all of these studies include only "volunteers," which raises the possibility that those who choose to participate may be unusually sensitive to price variation. Third, many of these pilots include a very small number of participants, and in at least one study a large fraction of those who started in the pilot dropped out before it was completed. Fourth, few of the pilot programs use full real-time pricing. A few use "critical peak pricing" mechanisms, and this may yield results similar to what we would get with full real-time pricing. For example, PG&E's voluntary tariff for customers with smart meters starts with the regular tariff price, except during "Smart High Price Periods," which are communicated to the customer in advance by telephone, Internet posting, or text messaging, and the price rises to 60 cents per kilowatt hour between 2 p.m. and 7 p.m. for a maximum of 15 days per summer season. Fifth, several of the pilots apply only one price to the treatment group, which makes it impossible to trace out the relevant demand functions without making very strong assumptions about the shape of the demand curves. Using several treatment groups requires a larger pilot study than has often been the case (see Aigner, 1985, regarding the need for multiple treatment groups). Finally, few of these studies use data on consumer responses along with electricity supply and metering cost data to perform a proper cost–benefit analysis. Indeed, I have not yet seen a recent study as well designed, or with a welfare analysis as carefully performed, as the Los Angeles experiments managed and analyzed by RAND during the 1970s (Mitchell and Acton 1980).

Despite these concerns, the available evidence does suggest a number of conclusions: First, consumers do respond to higher peak prices by reducing peak demand. Second, dynamic pricing with very high prices during critical periods generally leads to much larger prices responses than traditional time-of-use pricing with predetermined time periods and prices, which typically use much smaller price differences. Third, wide variation in price responsiveness is observed across studies, suggesting wide underlying variation in the attributes of households and pilot study conditions. Fourth, most if not all of the price response to higher peak period prices is to reduce peak demand rather than to shift from peak to off-peak demand. For example, a common reaction is to use less lighting, air conditioning, and refrigeration when prices are high—with no offsetting increase in electricity used at other times. However, the diffusion of plug-in vehicles or other technologies where time-of-use is a more important choice variable could yield very different results.[6] Fifth, technologies and information that make it easier for consumers to respond to high price signals lead to larger responses to any given price increase, although many of the reported results do not contain adequate information to estimate demand functions or to perform proper cost–benefit analyses.

Faruqui and Wood (2011) present a well-conceived "template" for what items should be included in a comprehensive cost–benefit analysis and present simulations for four "prototype" utilities. The aggregate benefit/cost ratios vary from 1.4 to 1.9 for the four simulations. The simulations are not based on real utilities nor complete data, but the hypothetical numbers are not unreasonable and the results are suggestive. Of course, cost–benefit analysis of universal smart meter installation and real-time pricing may also find that while the benefit/cost ratio is greater than 1.0 in the aggregate, it may not be beneficial to some significant number of individual customers. Borenstein (2007b) takes the wide variation in customer utilization attributes seriously, although his focus is on larger commercial and industrial customers, not residential customers. But the heterogeneity of the effects of smart meters and real-time pricing on residential and small commercial customers is an important issue that still needs to be addressed.

Some states that have mandated the installation of smart meters for all customers have found the decision to be controversial (for newspaper accounts, see Smith, 2009; Turkel, 2011a, b; Fehrenbacher, 2010; Baker, 2010). Some consumers have reacted negatively to increases in up-front distribution costs to pay for the smart grid enhancements. First, some customers with "unfavorable" consumption patterns—weighted toward times when prices are high—may see higher bills, rather than the lower bills they are being promised, compared to their billing under flat rates (Borenstein 2007b). Second, some smart meters have been deployed too quickly and have not worked properly. Third, with all of the data that these meters can collect, privacy advocates have raised concerns about

[6] "Storage space heating" allows off-peak electricity to be used to heat up special bricks or slabs or water tanks, which are then used as a source of warmth during on-peak hours. When storage space heating was introduced in Europe during the 1960s, it was consciously designed to shift demand to off-peak periods. It did such a good job that the peak shifted from day to night in England and northern Germany and the regulated prices no longer reflected the patterns of demand and cost. Steiner (1957) and Kahn (1970) discuss this "shifting peak" case theoretically. More generally, we should be reminded that we should not take our eyes off of the long-run equilibrium which may look very different from the short-run equilibrium—especially after technological change.

what data will be made widely available and how it may be used and protected. Finally, some public utility commissions and some utilities have done a poor job educating their customers and have rolled out their smart meter installation program too quickly. There are lessons to be learned about deployment strategy from these experiences.

Conclusions

The existing electricity distribution system is very old in many areas, and investments to replace key components will have to accelerate just to maintain the reliability of the system. These replacement programs should be consistent with longer-term strategies for modernizing the distribution system. However, there is a lot of uncertainty about the size of costs and benefits, and these costs and benefits vary across distribution feeders as well as customers and regions. The rate and direction of future technological change on both sides of the meter is also uncertain. The transition to smart grid technology is going to take years, and there are sure to be notable bumps along the way.

Accordingly, it seems to me that a sensible deployment strategy is to combine a long-run plan for rolling out smart-grid investments with well-designed pilots and experiments. Using randomized trials of smart grid technology and pricing, with a robust set of treatments and the "rest of the distribution grid" as the control, would allow much more confidence in estimates of demand response, meter and grid costs, reliability and power quality benefits, and other key outcomes. For example, Faruqui's (2011b) report on the peak-period price responses for 109 pilot programs displays responses between 5 to 50 percent of peak demand. An order-of-magnitude difference in measured price responses is just not good enough to do convincing cost–benefit analyses, especially with the other issues noted above. In turn, the information that emerges from these studies could be used to make mid-course corrections in the deployment strategy. Given the large investments contemplated in smart meters and complementary investments, along with the diverse uncertainties that we now face, rushing to deploy a particular set of technologies as quickly as possible is in my view a mistake.

Despite these reservations, the country is on a path to creating smarter transmission and distribution grids. Exactly how far and how fast we go remains quite uncertain, especially as the federal subsidies enacted in 2009 for promoting the smart grid come to an end.

■ *The views expressed here are my own and do not reflect the views of the Alfred P. Sloan Foundation, MIT, Exelon Corporation, Transcanada Corporation, or any other organization with which I am affiliated. I am an outside director of Exelon Corporation and of Transcanada Corporation. My other affiliations are identified on my CV at ⟨http://econ-www.mit.edu/faculty/pjoskow/cv⟩.*

References

Aigner, Dennis. 1985. "The Residential Time-of-Use Pricing Experiments: What Have We Learned?" Chap. 1 in *Social Experimentation,* edited by Jerry A. Hausman and David A. Wise. University of Chicago Press.

Allcott, Hunt. 2011. "Rethinking Real-Time Electricity Pricing." *Resources and Energy Economics* 33(4): 820–42.

Baker, David R. 2010. "Some PG&E Customers Want Choice on Smart Meters." *San Francisco Chronicle,* August 16, page D-1. http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2010/08/16/BUPK1EUELD.DTL.

Barbose, Galen, C. Goldman, R. Bharvirkar, N. Hopper, M. Ting, and B. Neenan. 2005. "Real Time Pricing as a Default or Optional Service for C&I Customers: A Comparative Analysis of Eight Case Studies." Working Paper LBNL-57661, Lawrence Berkeley National Laboratory.

Boiteux, Marcel. 1964a. "The Choice of Plant and Equipment for the Production of Electricity." In *Marginal Cost Pricing in Practice,* edited by J. R. Nelson, 199–214. Englewood, NJ: Prentice-Hall.

Boiteux, Marcel. 1964b. "Marginal Cost Pricing." In *Marginal Cost Pricing in Practice,* edited by J. R. Nelson, 51–58. Englewood, NJ: Prentice-Hall.

Boiteux, Marcel. 1964c. "Peak Load Pricing." *Journal of Business* 33(2): 157–79.

Borenstein, Severin. 2002. "The Trouble with Electricity Markets: Understanding California's Restructuring Disaster." *Journal of Economic Perspectives* 16(1): 191–211.

Borenstein, Severin. 2005. "The Long-Run Efficiency of Real-Time Electricity Pricing." *Energy Journal* 26(3): 93–116.

Borenstein, Severin. 2007a. "Customer Risk from Real-Time Retail Electricity Pricing: Bill Volatility and Hedgeability." *Energy Journal* 28(2): 111–30.

Borenstein, Severin. 2007b. "Wealth Transfers among Large Customers from Implementing Real-Time Retail Electricity Pricing." *Energy Journal* 28(2): 131–49.

Borenstein, Severin, and Stephen Holland. 2005. "On the Efficiency of Competitive Markets with Time-Invariant Retail Prices." *RAND Journal of Economics* 36(3): 469–93.

Browermaster, Dan. 2011. "Plug-in Vehicles and Their Impact." PowerPoint presentation. http://www.cacities.org/resource_files/29491.PGEPEVIntro(2011-01-18).pdf.

Clayton, Paul R. 2001. *Fundamentals of Electric Circuit Analysis.* Hoboken, NJ: John Wiley & Sons.

Committee on Assessment of Resource Needs for Fuel Cell and Hydrogen Technologies, National Research Council. 2010. *Transitions to Alternative Transportation Technologies—Plug-In Hybrid Electric Vehicles.* National Academies Press, Washington, D.C.

Electric Power Research Institute (EPRI). 2011a. "Estimating the Costs and Benefits of the Smart Grid: A Preliminary Estimate of the Investment Requirements and Resultant Benefits of a Fully Functioning Smart Grid." Technical Report 1022519, Electric Power Research Institute, Palo Alto, CA.

Electric Power Research Institute (EPRI). 2011b. "Transportation Electrification: A Technology Overview." Technical Report 1021334, Electric Power Research Institute, Palo Alto, CA.

Faruqui, Ahmad. 2011a. "Dynamic Pricing: The Top 10 Myths." The Brattle Group. April 7. http://www.brattle.com/_documents/UploadLibrary/Upload936.pdf.

Faruqui, Ahmad. 2011b. "109 and Counting: An Update on Dynamic Pricing Experiments." The Brattle Group, PowerPoint Presentation. July 12, 2011. Provided by the author.

Faruqui, Ahmed, and Sanem Sergici. 2010. "Household Response to Dynamic Pricing of Electricity: A Survey of 15 Experiments." *Journal of Regulatory Economics* 38(2): 193–225.

Faruqui, Ahmed, and Sanem Sergici. 2011. "Evaluation of Baltimore Gas and Electric's Smart Energy Pricing Plan." PowerPoint presentation, The Brattle Group. http://www.brattle.com/_documents/UploadLibrary/Upload940.pdf.

Faruqui, Ahmed, and Lisa Wood. 2011. "The Benefits of Smart Meters." PowerPoint presentation, Institute for Electric Efficiency.

Fehrenbacher, Katie. 2010. "PG&E's Smart Meter Report: A Case Study of Infrastructure over Customer." Gigaom, May 12. http://gigaom.com/cleantech/pges-smart-meter-report-a-case-study-of-infrastructure-over-customer/.

Hogan, William. 1992. "Contract Networks for Electric Power Transmission." *Journal of Regulatory Economics* 4(3): 211–42.

Joskow, Paul. 2001. "California's Electricity Crisis." *Oxford Review of Economic Policy* 17(3): 365–88.

Joskow, Paul. 2005. "Transmission Policy in the United States." *Utilities Policy* 13(2): 95–115.

Joskow, Paul. 2006. "Markets for Power in the United States: An Interim Assessment." *Energy Journal* 27(1): 1–36.

Joskow, Paul. 2011a. "Comparing the Costs of Intermittent and Dispatchable Generating Technologies." *American Economic Review* 101(3): 238–41.

Joskow, Paul. 2011b. "Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies." MIT CEEPR Working Paper (revised). http://econ-www.mit.edu/files/6317.

Joskow, Paul, and Jean Tirole. 2000. "Transmission Rights and Market Power on Electric Power Networks." *RAND Journal of Economics* 31(3): 450–87.

Joskow, Paul, and Jean Tirole. 2005. "Merchant Transmission Investment." *Journal of Industrial Economics* 53(2): 233–64.

Joskow, Paul, and Jean Tirole. 2007. "Reliability and Competitive Electricity Markets." *RAND Journal of Economics* 38(1): 68–84.

Kahn, Alfred. 1970. *Economic Principles.* Vol. 1 of *The Economics of Regulation: Principles and Institutions.* New York: John Wiley & Sons.

Mitchell, Bridger, and Jan Acton. 1980. "The Effects of Time-of-Use Rates in the Los Angeles Electricity Study." RAND, N-1533-DWP/HF.

Mitchell, Bridger, Willard Manning, and Jan Acton. 1978. *Peak-load Pricing: European Lessons for U.S. Energy Policy.* Cambridge, MA: Ballinger.

North American Electric Reliability Corporation (NERC). 2009. *Accommodating High Levels of Variable Generation: Special Report.*

PJM. 2007. *2006 State of the Market Report.* March 8. http://www.pjm.com/documents/reports/state-of-market-reports/2006-state-of-market-reports.aspx.

Power Engineering Society. 2006. "IEEE Benchmarking 2005 Results." PowerPoint Presentation.

Smith, Rebecca. 2009. "Smart Meter, Dumb Idea." *Wall Street Journal,* April 27, p. R5. http://online.wsj.com/article/SB124050416142448555.html.

Steiner, Peter O. 1957. "Peak Loads and Efficient Pricing." *Quarterly Journal of Economics* 71(4): 585–610.

St. John, Jeff. 2009. "8.3M Smart Meters and Counting in U.S." GreenTechGrid. July 17. http://www.greentechmedia.com/articles/read/8.3m-smart-meters-and-counting-in-united-states/.

Stoft, Stephen. 2002. *Power System Economics.* IEEE Press.

Taylor, Thomas N., Peter M. Schwartz, and James E. Cochell. 2005. "24/7 Hourly Response to Electricity Real-Time Pricing with up to Eight Summers of Experience." *Journal of Regulatory Economics* 27(3): 235–62.

Turkel, Tux. 2011a. "CMP: Smart Meter Bills Come with Huge Costs." *Portland Press Herald,* April 5, 2011. http://www.pressherald.com/news/cmp-smart-meter-bills-come-with-huge-costs_2011-04-05.html.

Turkel, Tux. 2011b. "PUC Allows 'Smart Meter' Opt-outs." *Portland Press Herald,* May 18, 2011. http://www.pressherald.com/news/puc-allows-smart-meter-opt-outs_2011-05-18.html.

Turvey, Ralph. 1968. "Peak Load Pricing." *Journal of Political Economy* 76(1): 101–13.

U.S. Federal Energy Regulatory Commission (FERC). 2011. *2010 Assessment of Demand Response and Advanced Metering.* Staff Report, February. http://www.ferc.gov/legal/staff-reports/2010-dr-report.pdf.

U.S. Energy Information Administration (EIA). 2010. *Annual Energy Review 2009.* Report No. DOE/EIA-0384(2009). Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2011a. *Annual Energy Outlook 2011,* Report No. DOE/EIA-0383(2011). Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2011b. "Advanced Electric Meter Installations Rising in Homes and Businesses." March 15. http://www.eia.gov/todayinenergy/detail.cfm?id=510.

Watkiss, Dan. 2011. "Divided Court Disconnects DOE Transmission Corridors." Bracewell & Giuliani Energy Legal Blog. February 8. http://www.energylegalblog.com/archives/2011/02/08/3483.

Wolak, Frank. 2006. "Residential Customer Response to Real-Time Pricing: The Anaheim Critical-Peak Pricing Experiment." CSEM Working Paper 151, Center for the Study of Energy Markets, University of California Energy Institute, Berkeley. http://www.ucei.berkeley.edu/PDF/csemwp151.pdf.

Wolak, Frank. 2010. "An Experimental Comparison of Critical Peak and Hourly Pricing: The PowerCents DC Program." Preliminary draft, March.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 49–66

Prospects for Nuclear Power

Lucas W. Davis

In September 2007, the U.S. Nuclear Regulatory Commission (NRC) received a license application for a proposed nuclear power reactor to be built in southern Texas. The application marked the first new license application in almost three decades. During the following year, the NRC received 16 license applications for a total of 24 proposed reactors. The time was right, so it seemed, for a nuclear power renaissance in the United States. Natural gas prices were at their highest level ever in real terms. The 2005 Energy Policy Act provided loan guarantees, production tax credits, and other subsidies for new nuclear plants. Many believed that the United States was close to enacting legislation that would limit emissions of carbon dioxide.

Then everything changed. Natural gas prices fell sharply in 2009. Legislation to limit carbon emissions stalled in Congress. The global recession slowed the growth of electricity demand. And in March 2011, an earthquake and tsunami knocked out power at the Fukushima Daiichi Nuclear Plant in northern Japan, causing partial meltdowns at the plant's three active reactors and large-scale releases of radioactive steam. Since 2009, only a single additional license application has been filed with the NRC. The project proposed for southern Texas has been canceled, and few of the applications pending with the NRC are moving forward. Fukushima has had perhaps an even stronger impact worldwide, leading Germany, Switzerland, and Italy to announce plans to phase out their nuclear power programs and causing China to suspend approvals for new reactors.

Nuclear power has long been controversial because of concerns about nuclear accidents, storage of spent fuel, and about how the spread of nuclear power might

■ *Lucas W. Davis is Assistant Professor at the Haas School of Business, University of California, Berkeley, California. His e-mail address is ⟨ldavis@haas.berkeley.edu⟩.*

doi=10.1257/jep.26.1.49

raise risks of the proliferation of nuclear weapons. These concerns are real and important. However, emphasizing these concerns implicitly suggests that unless these issues are taken into account, nuclear power would otherwise be cost effective compared to other forms of electricity generation. This implication is unwarranted. Throughout the history of nuclear power, a key challenge has been the high cost of construction for nuclear plants. Construction costs are high enough that it becomes difficult to make an economic argument for nuclear even before incorporating these external factors. This is particularly true in countries like the United States where recent technological advances have dramatically increased the availability of natural gas.

The chairman of one of the largest U.S. nuclear companies recently said that his company would not break ground on a new nuclear plant until the price of natural gas was more than double today's level and carbon emissions cost $25 per ton (Wald 2010). This comment summarizes the current economics of nuclear power pretty well. Yes, there is a certain confluence of factors that could make nuclear power a viable economic option. Otherwise, a nuclear power renaissance seems unlikely.

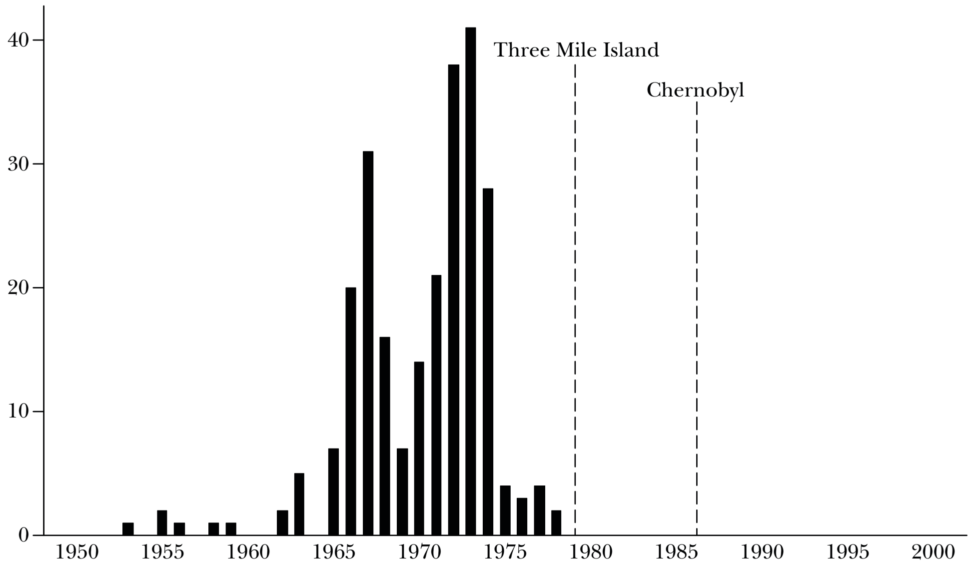
The First Boom and Bust

This recent ebb and flow in the nuclear power sector recalls a much larger boom and bust that occurred starting in the 1960s and 1970s. Figure 1 plots U.S. nuclear power reactor orders from 1950 to 2000. By 1974, there were 54 operating reactors in the United States with another 197 on order. This period was one of great enthusiasm for nuclear power. U.S. coal prices were at their highest level ever in real terms, and utilities were forecasting robust growth in electricity demand into the distant future.[1] The U.S. Atomic Energy Commission (1974) predicted that by the end of the twentieth century half of all U.S. electricity generation would come from nuclear power.

Instead, reactor orders fell precipitously after 1974. Over the next several years not only were new reactors not being ordered, but utilities began suspending construction on existing orders. Less than half of the reactors on order in 1974 were ever completed. Much has been written about the problems that faced the nuclear industry during this period (for example, Joskow and Yellin 1980; Joskow 1982; McCallion 1995). Part of the explanation is that concerns about safety and the environment began to take a more central role. In 1974, the Nuclear Regulatory Commission was created to replace the Atomic Energy Commission, which had previously been charged with both regulating and promoting nuclear

[1] For historic coal prices, see U.S. Department of Energy (2011a, table 7.9 "Coal Prices, 1949–2010"). Natural gas was much less important during the 1970s because modern combined cycle technology had not yet been widely introduced and because shortages associated with federal price controls on natural gas limited the availability of natural gas for electric generation.

Figure 1
U.S. Nuclear Power Reactor Orders



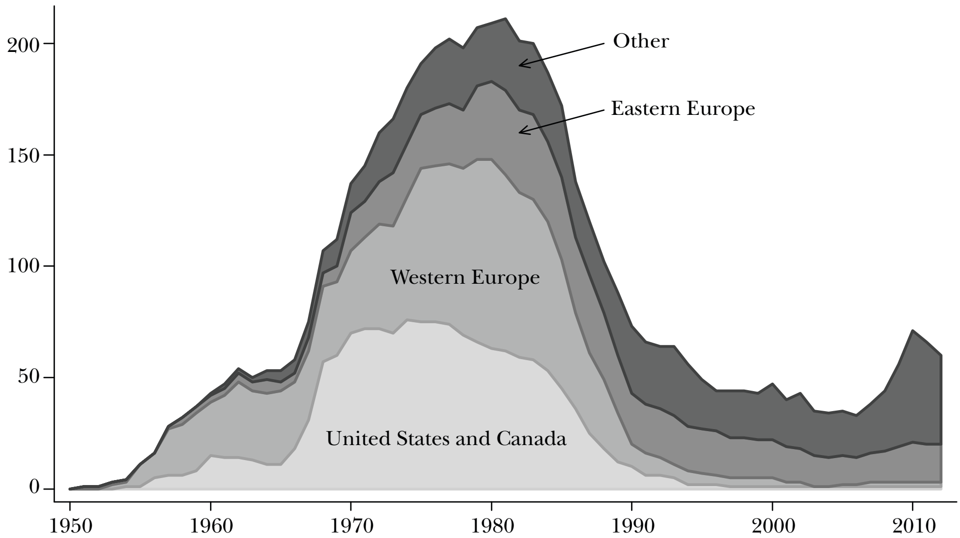
Source: Author based on data from U.S. Department of Energy (1997).

power—a combination of duties which many viewed to be in direct conflict. This new organization was to oversee the safety and security of all aspects of nuclear power, including the initial licensing of reactors, the handling of radioactive materials, and the storage and disposal of spent fuels. Beginning in the 1970s, it also became more difficult to site nuclear power plants. Communities began challenging nuclear power projects in federal and state courts, leading to extended construction delays and changing public attitudes about nuclear power.

Utility regulation also experienced structural change at this time. During the 1950s and 1960s, economies of scale, decreasing commodity costs, and relatively low inflation led to steady decreases in the nominal cost of electricity. Public utility commissions and consumers were pacified with prices that remained essentially the same in nominal terms year after year. Joskow (1974) explains that inflation in the early 1970s, "wreaked havoc on this process that appeared to function so smoothly before . . . and most major firms found that they had to raise prices (some for the first time in 25 years) and trigger formal regulatory reviews." These reviews led to increased scrutiny of utilities' capital expenditures, and in particular, investments in nuclear plants.

Then in March 1979, one of the reactors at the Three Mile Island plant in Pennsylvania suffered a partial core meltdown. Although not a single person was injured, the accident intensified public concerns about nuclear safety. The combination of severe public concern about the risk of nuclear accidents and

Figure 2
Nuclear Reactors under Construction Worldwide



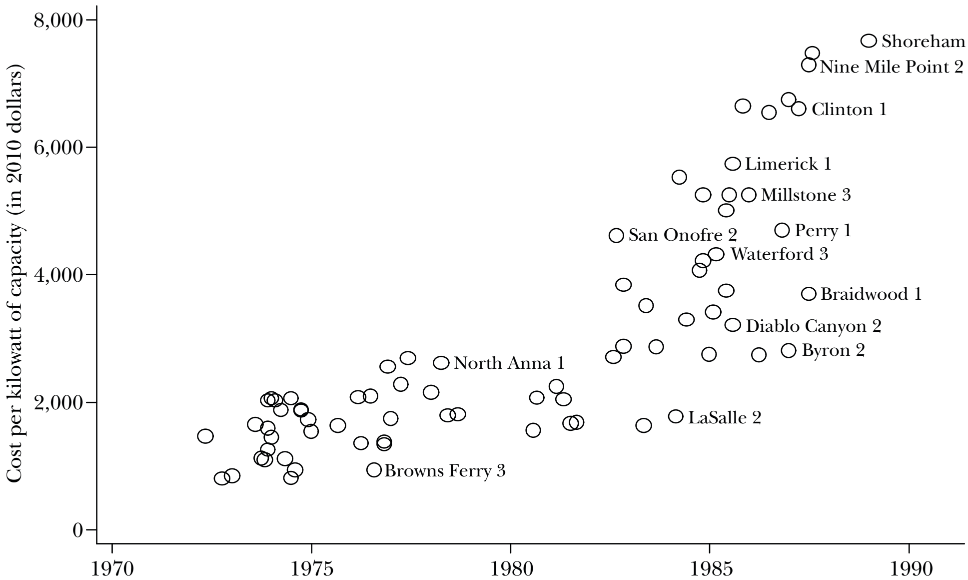
Source: Author based on data from International Atomic Energy Agency (2011).

escalating construction costs put nuclear projects in an extremely vulnerable position. By the time the Chernobyl disaster occurred in April 1986, expansion of U.S. nuclear power had largely halted. Today in the United States, there are a total of 104 nuclear power reactors at 65 sites, accounting for 20 percent of U.S. electricity generation (U.S. Department of Energy 2011a, table 8.2a).[2] All of these reactors were ordered prior to 1974.

Nuclear reactor construction outside the United States followed a similar pattern with a substantial boom in the 1960s and 1970s, followed by a long period of decline, shown in Figure 2. In addition to the United States, the other large-scale early adopters of nuclear power were the United Kingdom, France, Germany, Canada, and Japan. By the 1990s, construction had moved to Eastern Europe and in particular Russia and the Ukraine. The increase in construction 2008–2010 comes primarily from China, which today has more reactors under construction than any other country.

[2] U.S. Department of Energy (2011a, table 8.2a) also reports that U.S. net generation of electricity in 2010 includes coal (45 percent), natural gas (24 percent), nuclear (20 percent), hydroelectric power (7 percent), and wind and other renewables (4 percent). In Davis and Wolfram (2011), my coauthor and I examine in detail operating performance at U.S. nuclear plants.

Figure 3
Construction Costs for U.S. Nuclear Reactors by Year of Completion



Source: U.S. DOE (1986), table 4.
Notes: Figure 3 plots "overnight" construction costs for selected U.S. nuclear power plants from the U.S. Department of Energy (1986). The figure includes *predicted* costs from the same source for a handful of reactors that were under construction but not yet in operation in 1986.

Historical Construction Costs

Nuclear power plants are characterized by high construction costs and relatively low operating costs. Later in the paper, I present estimates of "levelized costs," which facilitate comparisons with other generating technologies with different cost profiles over time. It is worth starting with construction costs, however, because they represent a large share of the total cost of nuclear power.

Nuclear power plants are enormous facilities with high construction costs. The sheer scale of commercial-sized nuclear reactors means that most components must be specially designed and constructed, often with few potential suppliers worldwide. These components are then assembled on site, and structures are constructed to house the assembled components. All stages of design, construction, assembly, and testing require highly-skilled, highly-specialized engineers. Differences in reactor design and site-specific factors have historically meant that there was little scope for spreading design and production costs across multiple projects.

Figure 3 plots "overnight" construction costs for selected U.S. nuclear reactors from the U.S. Department of Energy (1986). The overnight cost is the hypothetical cost of a plant if it could be built instantly and thus excludes financing and other costs incurred during plant construction. Costs are reported in year 2010 dollars

Table 1
U.S. Nuclear Reactor Orders and Construction Time

Decade	*Number of reactors ordered*	*Percent eventually completed*	*Construction time (in years) for completed reactors*		
			Average	*Minimum*	*Maximum*
1950s	6	100%	4.5	3	7
1960s	88	89%	8.6	3	22
1970s	155	25%	14.1	8	26

Source: Author's tabulations based on U.S. Department of Energy (1997). Construction time is calculated as the difference in years between when a reactor is ordered and when it begins commercial operation.

per kilowatt of capacity. The figure reveals a pronounced increase in construction costs, particularly for plants completed during the 1980s.[3]

Plants also kept taking longer and longer to build. As shown in Table 1, reactors ordered during the 1950s took on average about five years to build, whereas reactors ordered during the 1970s took on average 14 years. Most studies attribute this increase in construction time to a rapidly evolving regulatory process. A joke in the industry was that a reactor vessel could not be shipped until the total weight of all required paperwork had equaled the weight of the reactor vessel itself. Regulation also contributed directly to construction costs. The Nuclear Regulatory Commission implemented revised safety codes and inspection requirements leading in several cases to extensive reactor redesigns (Cox and Gilbert 1991; McCallion 1995).

An interesting point of comparison is France. After the United States, France has more nuclear reactors than any other country, and 75 percent of electricity generation in France comes from nuclear power. Grubler (2010) finds that 58 reactors in France's main nuclear program were constructed at an average cost that increased over time from $1,000 per kilowatt of capacity in the 1970s to $2,300 in the 1990s. The cost escalation is less severe than is observed in the United States, but still somewhat surprising. As I discuss later, in many ways the French nuclear program was the ideal setting for encouraging learning-by-doing, so one might have expected costs to decrease over time.

Financing Risks

The long period of time required for construction means that the cost of capital is a critical parameter for evaluating the viability of nuclear power. Even for a low

[3] See Koomey and Hultman (2007) for a more recent study of U.S. nuclear construction costs, and Mooz (1978), Komanoff (1981), and Zimmerman (1982) for studies of the earlier period. Joskow and Rose (1985) examine increases in construction costs for coal plants during the 1960s and 1970s finding significant cost increases associated with measurable environmental-related technologies such as scrubbers and cooling towers, as well as a large increase in residual real costs that they attribute to changes in environmental regulation and to an unexplained decline in construction productivity.

cost of capital, the extended construction period imposes financing costs that are a substantial part of total project costs. However, nuclear projects typically face a cost of capital well above the risk-free rate. These large-scale projects have a historically high risk of default. The high cost of capital that they face reflects the number of risks that threaten the profitability, and even viability, of a nuclear project.

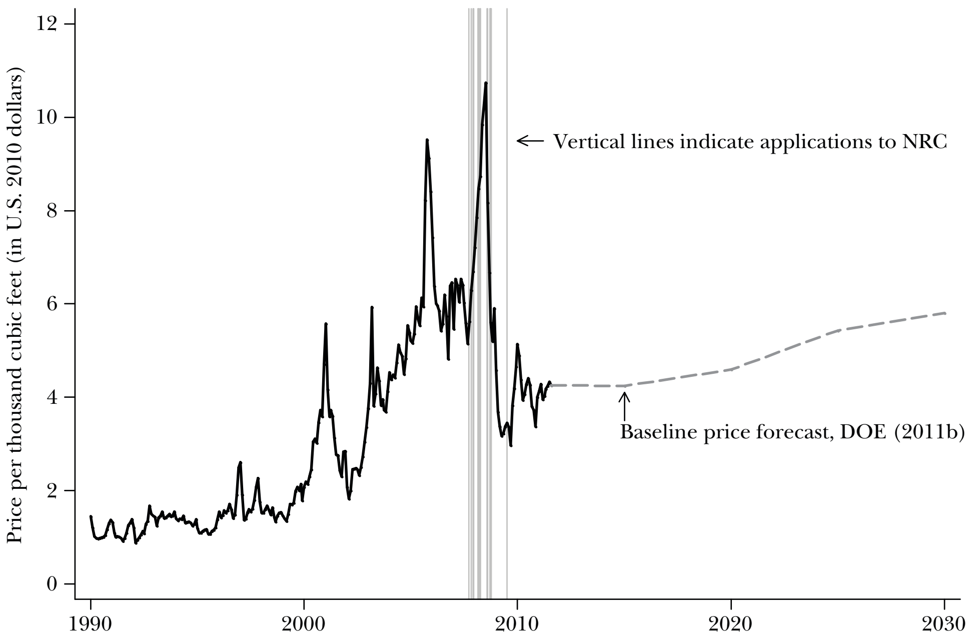
More so than in most other investments, nuclear power plants face substantial regulatory risk. The Nuclear Regulatory Commission has recently adopted several new procedures intended to streamline the regulatory process. These reforms include pre-approving standard reactor designs, an early site permitting process, and combining construction and operating licenses which previously were applied for separately. It remains to be seen how these procedures will work in practice. Regulatory approval is also required at the state and local level, and it can be a real constraint on plants. For example, in 1989 New York Governor Mario Cuomo and the Long Island Lighting Company closed the Shoreham Nuclear Power Plant over long-standing concerns about how nearby residents would be evacuated in the event of an emergency. The plant was 100 percent completed and had been connected to the grid, yet was never used to produce a single kilowatt hour of commercial electricity.

Nuclear power is also sensitive to federal energy policy. The enthusiasm for nuclear power in 2007 and 2008 was driven in part by the prospect of a federal cap on carbon emissions, and so when the key legislative vehicle (H.R. 2454, the "Waxman–Markey bill") stalled in the U.S. Senate in 2009, it was a significant blow to the economic viability of new nuclear plants. In the last few years, the Obama administration and some members of Congress have voiced support for a federal "clean energy standard" under which a proportion of total electricity generation would be required to come from sources that do not generate carbon emissions. Such a policy could be a considerable boost for nuclear power, but the exact form of such legislation, or how likely its adoption would be, is unclear.

Investors in nuclear power also face the risk that fossil fuel prices could decrease. In the United States, natural gas prices typically determine the marginal cost of electricity, so a decrease in natural gas prices reduces profits for nuclear plants that sell power in wholesale electricity markets. Global availability of natural gas has increased dramatically in recent years with improvements in horizontal drilling and hydraulic fracturing technology. Natural gas producers have long known that shale and other rock deposits contain large amounts of natural gas. It was not until recently, however, that these resources could be accessed at reasonably low cost.

Figure 4 plots U.S. natural gas prices from 1990 to 2011 and a price forecast through 2030. During the long period of relatively low natural gas prices, there was not a single new nuclear plant ordered in the United States, and the surge in orders in 2007 and 2008 came at the same time that U.S. natural gas prices reached their highest level ever in real terms. The baseline forecast from U.S. Department of Energy (2011b) predicts that U.S. natural gas prices will remain under $5 per thousand cubic feet through 2022. If true, this is a significant challenge for nuclear power.

Figure 4
U.S. Natural Gas Prices and Applications to the U.S. Nuclear Regulatory Commission (NRC)



Sources: Author based on data from U.S. Department of Energy (2011a, b) and U.S. Nuclear Regulatory Commission (2011).

Finally, investments in nuclear power face considerable technology risk. Over the 40-plus year lifetime of a nuclear plant, the available sources of electricity generation could change considerably. An alternative, lower-cost technology could come along, or perhaps a technology that is known today such as wind or solar could quickly become more cost effective. An alternative technology for carbon abatement could become practical, like some form of carbon capture and storage, which would render moot one of the advantages of nuclear power. New energy efficiency technologies might reduce electricity demand.

Recent International Experience

More recent evidence on construction costs comes from nuclear reactors currently being built in Olkiluoto, Finland, and Flamanville, France. Much has been written about these reactors because they are the first new reactors to be built in Europe in many years, and because they use a "next generation" design that incorporates several new safety features into a reactor design that is widely used around the world. Construction in Finland began in 2005 and was expected to

Table 2
Power Plant Construction Costs Excluding Financing for Nuclear, Coal, and Natural Gas

	Cost per kilowatt of capacity (in year 2010 dollars)		
Source	*Nuclear*	*Coal*	*Natural gas*
MIT (2009)	$4,200	$2,400	$ 900
U.S. Department of Energy (2010)	$5,300	$2,800	$1,000

be completed in 2009 at cost of about $2,800 per kilowatt of capacity. A series of problems and delays have now pushed operations back to 2013, and costs are now estimated to be about twice the original estimate. Similarly, construction in France began in 2007 and the reactor was expected to be completed by 2011 at a cost of $2,900 per kilowatt. Completion has now been pushed back to 2014 and the project is reported to be 50 percent over budget.

These experiences provide a reminder about problems that can occur during reactor construction, particularly given the lack of recent construction experience. Both projects were delayed substantially when government safety inspectors found problems. In Finland, the concrete foundation of the reactor building was found to be too porous. In France, inspectors found cracks in the concrete foundation and steel reinforcements in the wrong places. Project managers have been blamed in both projects for hiring inexperienced contractors and for providing insufficient oversight.

Construction costs have tended to be lower elsewhere. Du and Parsons (2009) report a mean overnight cost of $3,100 per kilowatt from five reactors completed in Korea and Japan between 2004 and 2006. Construction costs from plants recently completed in China are reported to be even lower and an important area for future research is to examine these costs in detail.

Several studies have attempted to synthesize this recent international construction experience with historical U.S. data and engineering studies to estimate current construction costs for the United States. Table 2 reports estimates of overnight construction cost from two such studies. MIT (2009) estimates $4,200 per kilowatt of capacity for nuclear, compared to $2,400 and $900 per kilowatt of capacity for coal and natural gas. U.S. Department of Energy (2010) predicts somewhat higher costs particularly for nuclear plants, citing increased prices for plant components and key commodities and arguing that costs will be driven up by the fact that only a limited set of construction firms have the ability to complete a project of this scale. Both studies were completed prior to Fukushima and thus do not incorporate any cost increases due to recently elevated regulatory scrutiny.

Adding financing costs to these estimates implies that a typical two-reactor 2,000 megawatt plant could cost more than $12 billion. The long period of time since nuclear power plants were constructed in the United States means that the relevant experience that had been accumulated by companies involved with nuclear engineering and plant construction has atrophied substantially (Joskow and Parsons 2009). There is some scope for importing nuclear engineers and other professionals who have worked on more recent nuclear projects in other countries, but the overall level of nuclear construction activity worldwide over the last 20 years has been so low that the available global talent is limited. Moreover, the supply of nuclear plant components is now more limited than it was during the first wave of nuclear power plant construction. For example, there is currently only one facility in the world that can produce the nuclear-grade heavy-steel reactor vessel needed for a boiling water reactor, and there is currently a long waiting period for these forgings and for other key nuclear components (Ives, McCabe, and Gilmartin 2010).

These construction cost estimates contain considerable uncertainty, which is itself a barrier to investment. Pindyck (1993) uses a model of irreversible investment to illustrate how uncertainty over the prices of construction inputs and over government regulation affecting construction costs can lead investors to delay investment on nuclear projects. One of the economic arguments made in support of the subsidies for new nuclear plants in the 2005 Energy Policy Act was that they would help resolve this uncertainty about construction costs.

Levelized Cost Estimates

The total cost of producing electricity depends both on construction costs and on operations and maintenance expenditures, including fuel. These variable costs tend to be low for nuclear, potentially offsetting the higher cost of construction. Table 3 reports "levelized" costs for electricity generated in the United States from nuclear, coal, and natural gas, the three primary forms of baseload electricity generation. These estimates are based on a cash flow model developed in an ongoing series of studies at the Massachusetts Institute of Technology (MIT 2003; MIT 2009; Joskow 2006; Du and Parsons 2009; Joskow and Parsons 2009). For these estimates, all costs including construction, operation, maintenance and fuel are calculated and discounted back to the present using an assumed cost of capital. This total cost is then "levelized" over the lifetime of a plant in constant dollars to yield the long-run average cost of producing a kilowatt hour of electricity. This is equivalent to the real price per kilowatt hour that the plant would need to receive over its lifetime in order to break even.

Under the baseline assumptions, nuclear is not competitive with either coal or natural gas. The first row of Table 3 reports the base case estimates reported in MIT (2009). The levelized cost of nuclear power is 8.7 cents per kilowatt hour, compared to 6.5 cents for coal and 6.7 cents for natural gas. This gap widens in the second row after updating these estimates to reflect higher construction cost estimates from U.S. Department of Energy (2010). The third row updates the estimates to reflect changes in fuel prices since 2009. Uranium prices have increased modestly, but fuel expenditures represent a relatively small proportion of the total cost of nuclear power, and at this higher price, even after including costs for conversion,

Table 3
Levelized Cost Comparison for Electricity Generation

	Levelized cost in cents per kWh		
Source	*Nuclear*	*Coal*	*Natural gas*
MIT (2009) baseline	8.7	6.5	6.7
Updated construction costs	10.4	7.0	6.9
Updated construction costs and fuel prices	10.5	7.4	5.2
With carbon tax of $25 per ton CO_2	10.5	9.6	6.2

Source: These calculations follow MIT (2009) except where indicated in the row headings.
Notes: All costs are reported in 2010 cents per kilowatt hour. Row 1 reports the base case estimates reported in MIT (2009), table 1. The cost estimates reported in row 2 incorporate updated construction cost estimates from U.S. Department of Energy (2010). Row 3, in addition, updates fuel prices to reflect the most recent available prices for uranium, coal, and natural gas reported in U.S. DOE (2011a). Finally, row 4 continues to incorporate updated construction costs and fuel prices and, in addition, adds a carbon tax of $25 per ton of carbon dioxide.

enrichment, and fuel fabrication, nuclear fuel costs are still less than one cent per kilowatt hour. Moreover, the medium- to long-run supply of uranium is highly elastic, with substantial known reserves worldwide with a cost of recovery below current uranium prices (MIT 2003, Appendix 5.E; OECD 2009). Fossil fuel prices are extremely important for the prospects for nuclear power, and the cost estimates in the third row reflect somewhat higher coal prices but also considerably *lower* natural gas prices. With these updated prices, the levelized cost of electricity from natural gas is just above 5 cents per kilowatt hour, compared to more than 10 cents per kilowatt hour for nuclear.

These estimates follow the MIT studies in applying a somewhat higher cost of capital to nuclear power. As discussed earlier, this reflects the high risk of default and numerous forms of risk faced with nuclear projects. It is worth noting, however, that even without this risk premium, nuclear still has higher levelized cost than coal or natural gas. The model also assumes a 40-year lifetime for nuclear, coal, and natural gas plants. Over half of U.S. nuclear plants have received license extensions to 60 years. Incorporating a longer lifetime into the model makes nuclear look better, but not by very much. The increased net revenue is far in the future so with discounting there is only a modest decrease in levelized costs. Moreover, coal and natural gas plants are also tending to be used for more than 40 years, and one would want to incorporate those longer lifetimes as well.

It is important to emphasize that these levelized cost estimates depend on a series of empirical assumptions, many of which can be only partially verified. Perhaps most importantly, alternative assumptions about nuclear construction costs or natural gas prices can begin to change the outlook considerably. Moreover, these cost estimates are for the United States and may not easily generalize to other countries. Construction costs vary substantially across countries due to differences in the cost of labor and other inputs, as well as differences in the regulatory environment.

Another key factor is the availability of natural gas. Global capacity to transport liquefied natural gas is increasing rapidly but is still insufficient to have eliminated price differences across countries.

Incorporating Externalities

Levelized cost estimates are typically designed to reflect the *private* costs of investing in different forms of electricity generation. Thus they provide a basis for determining whether different types of plants *will* be built, but not necessarily for determining whether different types of plants *should* be built.

The fourth row in Table 3 incorporates a tax of $25 per ton of carbon dioxide. As a point of comparison, the Federal Interagency Working Group (2010) adopts a central social cost of carbon dioxide of $25 for 2015. Under this scenario, nuclear continues to have the highest levelized cost. The levelized cost of coal increases by 2 cents per kilowatt hour, but the levelized cost of natural gas increases by only about 1 cent, not nearly enough to close the gap between nuclear and natural gas. Moreover, this static comparison based on current fuel prices ignores that coal and natural gas prices would likely fall in response to a carbon tax. For both coal and natural gas there is a range of different sources available, much of which has a marginal cost of extraction below current prices.

Fossil fuel plants also emit large amounts of local and regional pollutants. Muller, Mendelsohn, and Nordhaus (2011) calculate that the external costs from sulfur dioxide, nitrogen oxides, and particulates average 3.5 cents per kilowatt hour for coal, but only 0.1 cents per kilowatt hour for natural gas. Thus, incorporating the external costs of these pollutants improves the prospects considerably for nuclear power versus coal, but does little to close the gap versus natural gas. A comprehensive welfare analysis would also incorporate the negative production externalities from coal and natural gas. Perhaps most importantly, recent increases in shale gas production have raised environmental concerns about water consumption and contamination of drinking water. These costs are not yet well understood. However, the levelized cost estimates give some sense of how large these externalities would need to be in order to make nuclear power the low-cost option.

There are also external costs associated with nuclear power. Included in these levelized cost estimates is a spent fuel waste fee of 0.1 cents per kilowatt hour. Since 1983, the Department of Energy has collected this fee from U.S. nuclear reactors, intended eventually to finance a centralized storage facility for spent nuclear fuel. Currently, most spent nuclear fuel is stored on-site in spent fuel pools and dry casks. A comprehensive welfare analysis would need to include both the private and external costs of this on-site storage. MIT (2010) and U.S. Nuclear Regulatory Commission (2011) discuss details of the nuclear fuel cycle.

Considerably harder to quantify are the risks from nuclear accidents. Since 1957, the Price–Anderson Act has indemnified U.S. nuclear plant operators from accident liability above a certain cap, currently $12 billion. A Fukushima-type accident in the United States could easily cause damages well above this cap. It is too early to measure the long-term external costs of Fukushima, but an early study estimates that radioactive contamination could cause 1,000 total cancer deaths (von Hippel 2011). As a point of comparison, cancer deaths from Chernobyl are estimated to be approximately 14,000 (Cardis et al. 2006). In addition to cancer deaths, one would want to incorporate the costs from other health outcomes, as well as the pecuniary and psychological costs associated with relocating people living near the accident site.

Perhaps hardest of all to measure are the risks associated with the proliferation of nuclear weapons. This could come through the misuse of nuclear facilities to produce weapons materials, or from a "dirty bomb" in which stolen radioactive materials from any source are dispersed using conventional explosives. These risks are particularly acute in countries like France, the United Kingdom, and Japan that have facilities for reprocessing nuclear waste. MIT (2003) and MIT (2010) discuss these issues.

Incorporating the external costs of nuclear power would further increase the gap between the levelized costs of nuclear and natural gas. An important priority for future work is to refine measures of these external costs and incorporate them explicitly into levelized cost analyses. However, given current market conditions in the United States, it becomes difficult to make an economic argument for nuclear power regardless of the magnitude of these external costs. The first challenge continues to be construction costs, which are high enough that nuclear power struggles to compete with natural gas even if one ignores these external costs completely.

Learning-By-Doing

What would it take to reduce nuclear construction costs? One possibility is learning-by-doing. In 2004, the Senior Vice President of the Nuclear Energy Institute testified in front of the U.S. Senate that nuclear construction costs would decrease by 20–30 percent after the first few plants (Fertel 2004). In part on the basis of this testimony, the 2005 Energy Policy Act was drafted to include loan guarantees, production tax credits, and other subsidies for new nuclear plants. If learning-by-doing could push construction costs down, this could change the equation considerably for nuclear power. A substantial literature in economics indicates that learning-by-doing matters in a variety of markets (Alchian 1963; Joskow and Rose 1985; Irwin and Klenow 1994; Benkard 2000; Thornton and Thompson 2001; Kellogg 2011), and several studies have examined learning-by-doing in the construction of nuclear power plants.

Recall that the time pattern of construction costs in Figure 3 did not provide any immediate evidence of learning-by-doing. Instead, construction costs tended to *increase* considerably over time. Several studies have nonetheless attempted to disentangle learning-by-doing from industrywide factors that were changing over time. Using data from the early nuclear builds, both Mooz (1978) and Komanoff (1981) find evidence of modest amounts of learning-by-doing in nuclear plant construction that accrue to the construction company in charge of the project, but no evidence of industrywide learning-by-doing. Zimmerman (1982) also finds learning-by-doing for the construction company and some evidence of spillovers across companies. Using a longer panel, McCabe (1996) finds evidence of learning-by-doing for both the construction company and the utility managing the project, but does not test for industrywide learning.

Learning-by-doing is important for the prospects of nuclear power because it provides a plausible mechanism by which nuclear construction costs could decrease below the levels reported in Table 2. This is true regardless of whether or not this learning-by-doing is privately captured. Who captures the learning-by-doing is important, however, for government policy. The economic argument for an industry-specific subsidy hinges on there being learning-by-doing that is not captured by individual companies. If learning is fully appropriable, then firms face efficient incentives for investment and no government intervention is necessary. In addition, while there is almost certainly some industrywide learning-by-doing in nuclear, there is also likely to be learning-by-doing in emerging energy technologies such as wind, solar, and biomass. When there are a number of competing alternatives, as in electricity generation, many economists favor broad-based subsidies that do not single out individual technologies (Schmalensee 1980).

Tied up in this discussion is a key tradeoff between innovation and standardization. On the one hand, it is important to continue allowing for new and better reactor designs with enhanced features for reliability and safety. On the other hand, frequent redesigns make it harder to spread engineering costs across projects. The first wave of U.S. reactors were manufactured by four different companies—Westinghouse, General Electric, Combustion Engineering, and Babcock & Wilcox—each with several different designs. At the time, such differences were inevitable. The United States led the way in the development of commercial nuclear reactors and the technology was evolving rapidly. Still, this diversity of designs provides a possible explanation for the lack of immediate evidence of learning-by-doing (Lester and McCabe 1993).

France offers a useful comparison on this point. Development of nuclear power in France began later and with much less design variation. When Electricité de France began seriously building reactors in the 1970s, it adopted a single design for all of its reactors. With one exception, all nuclear power reactors currently in operation in France are of exactly this same design (International Atomic Energy Agency 2011). In addition, Electricité de France has long enjoyed a high degree of regulatory stability due to its close relationship with the French National Safety Authority and broad public support for nuclear power. Given this high degree of standardization, the apparent cost *escalation* in French construction costs is particularly striking.

Some within the U.S. nuclear industry claim that the United States is headed more toward the French model. For example, Michael Wallace, chairman of a major nuclear power company predicted a couple of years ago (as quoted in Kanter 2009) that new reactors would be standardized down to "the carpeting and wallpaper." Perhaps the industry will quickly coalesce around a very small number of reactor designs, but this is not obvious based on applications received to date by the Nuclear Regulatory Commission. Among the 17 applications that have been received, there is a mix of both pressurized water reactors and boiling water reactors, manufactured by five different reactor manufacturers (Areva, Westinghouse, Mitsubishi, GE-Hitachi, and GE). At a minimum, it seems clear that the French approach of supporting a single reactor design is not going to be adopted in the United States.

Conclusion

Nuclear power continues to generate enthusiasm based on its potential to reduce greenhouse gas emissions. A single pound of reactor-grade uranium oxide produces as much electricity as over 16,000 pounds of coal—enough to meet the needs of the average U.S. household for more than a year.[4] While burning 16,000 pounds of coal generates thousands of pounds of carbon dioxide, sulfur dioxide, and nitrogen oxides—nuclear power is virtually emissions-free.

Nuclear power, however, is not without challenges. Fukushima has brought to the forefront ongoing concerns about nuclear accidents and the handling and storage of spent fuel. These external costs are *in addition* to substantial private costs. In 1942, with a shoestring budget in an abandoned squash court at the University of Chicago, Enrico Fermi demonstrated that electricity could be generated using a self-sustaining nuclear reaction. Seventy years later the industry is still trying to demonstrate how this can be scaled up cheaply enough to compete with coal and natural gas.

■ *I am thankful to David Autor, Severin Borenstein, Chad Jones, John List, John Parsons, Timothy Taylor, Catherine Wolfram, and seminar participants at the University of Tennessee, Brookings Institution, University of California Energy Institute, Berkeley Energy and Resources Collaborative Symposium, California Public Utilities Commission, and the University of Chicago for helpful comments.*

[4] This is my own back-of-the-envelope calculation based on U.S. Department of Energy (2011a), table 8.2a "Electricity Net Generation," table 8.5a "Consumption of Combustible Fuels," table 8.9 "Electricity End Use," table 9.3 "Uranium Overview," and table 12.7b "Emissions from Energy Consumption for Electricity Generation."

References

Alchian, Armen. 1963. "Reliability of Progress Curves in Airframe Production." *Econometrica* 31(4): 679–94.

Benkard, C. Lanier. 2000. "Learning and Forgetting: The Dynamics of Aircraft Production." *American Economic Review* 90(4): 1034–54.

Cardis, Elisabeth, et al. 2006. "Estimates of the Cancer Burden in Europe from Radioactive Fallout from the Chernobyl Accident." *International Journal of Cancer* 119(6): 1224–35.

Cox, Alan J., and Richard J. Gilbert. 1991. "An Economic Evaluation of the Costs and Benefits of Diablo Canyon." In *Regulatory Choices: A Perspective on Developments in Energy Policy,* edited by Richard J. Gilvert, 260–89. Berkeley: University of California Press.

Davis, Lucas W., and Catherine Wolfram. 2011. "Deregulation, Consolidation, and Efficiency: Evidence from U.S. Nuclear Power." EI@Haas WP 217, Energy Institute at Haas.

Du, Yangbo, and John E. Parsons. 2009. "Update on the Cost of Nuclear Power." MIT Center for Energy and Environmental Policy Research Working Paper 09-004.

Federal Interagency Working Group. 2010 "Technical Support Document: Social Cost of Carbon for Regulatory Impact Analysis: Under Executive Order 12866." http://www.epa.gov/otaq/climate/regulations/scc-tsd.pdf.

Fertel, Marvin S. (Senior Vice President and Chief Nuclear Officer of the Nuclear Energy Institute). 2004."Testimony before the Senate Energy and Natural Resources Committee, U.S. Senate." March 4. http://www.nei.org/publicpolicy/congressional testimony/energysubcmtefertelextended/.

Grubler, Arnulf. 2010. "The Costs of the French Nuclear Scale-Up: A Case of Negative Learning by Doing." *Energy Policy* 38(9): 5174–88.

International Atomic Energy Agency. 2011. *Nuclear Power Reactors in the World.* 2011 Edition. Vienna: IAEA.

Irwin, Douglas A., and Peter J. Klenow. 1994. "Learning-by-Doing Spillovers in the Semiconductor Industry." *Journal of Political Economy* 102(6): 1200–27.

Ives, Nathan, Steve McCabe, and Gary Gilmartin. 2010 "Nuclear Renaissance and the Global Supply Chain: Avoiding Pitfalls, Realizing Benefits." *Public Utilities Fortnightly,* December 1.

Joskow, Paul L. 1974. "Inflation and Environmental Concern: Structural Change in the Process of Public Utility Price Regulation." *Journal and Law and Economics* 17(2): 291–327.

Joskow, Paul L. 1982. "Problems and Prospects for Nuclear Energy in the United States." In *Energy, Economics and the Environment,* edited by Gregory A. Daneke, 231–54. Lexington: D.C. Heath and Company.

Joskow, Paul L. 2006. "The Future of Nuclear Power in the United States: Economic and Regulatory Challenges." MIT CEEPR Working Paper 06-019, Massachusetts Institute of Technology Center for Energy and Environmental Policy Research.

Joskow, Paul L., and John E. Parsons. 2009. "The Economic Future of Nuclear Power." *Daedalus* 138(4): 45–59.

Joskow, Paul L., and Nancy L. Rose. 1985. "The Effects of Technological Change, Experience, and Environmental Regulation on the Construction Cost of Coal-Burning Generating Units." *RAND Journal of Economics* 16(1): 1–17.

Joskow, Paul L., and Joel Yellin. 1980. "Siting Nuclear Power Plants." *Virginia Journal of Natural Resource Law* 1(1): 1–67.

Kanter, James. 2009. "In Finland, Nuclear Renaissance Runs into Trouble." *New York Times,* May 28. http://www.nytimes.com/2009/05/29/business/energy-environment/29nuke.html?pagewanted=all.

Kellogg, Ryan. 2011. "Learning by Drilling: Inter-Firm Learning and Relationship Persistence in the Texas Oilpatch." *Quarterly Journal of Economics* 126(4): 1961–2004.

Komanoff, Charles. 1981. *Power Plant Escalation: Nuclear and Coal Capital Costs, Regulation, and Economics.* New York: Van Nostrand Reinhold Company.

Koomey, Jonathan, and Nathan E. Hultman. 2007. "A Reactor-Level Analysis of Busbar Costs for U.S. Nuclear Plants, 1970–2005." *Energy Policy* 35(11): 5630–42.

Lester, Richard K., and Mark J. McCabe. 1993. "The Effect of Industrial Structure on Learning by Doing in Nuclear Power Plant Operation." *RAND Journal of Economics* 24(3): 418–38.

McCabe, Mark J. 1996. "Principals, Agents, and the Learning Curve: The Case of Steam-Electric Power Plant Design and Construction." *Journal of Industrial Economics* 44(4): 357–75.

McCallion, Kenneth. 1995. *Shoreham and the Rise and Fall of the Nuclear Power Industry.* Praeger, Westport.

MIT. 2003. *The Future of Nuclear Power.* An Interdisciplinary MIT Study. MIT Energy Initiative, Massachusetts Institute of Technology.

MIT. 2009. "*Update of the MIT 2003 Future of Nuclear Power.* An Interdisciplinary MIT Study."

MIT Energy Initiative, Massachusetts Institute of Technology.

MIT. 2010. *The Future of the Nuclear Fuel Cycle.* An Interdisciplinary MIT Study. MIT Energy Initiative, Massachusetts Institute of Technology.

Mooz, William E. 1978. *Cost Analysis of Light Water Reactor Power Plants.* RAND Report, Prepared for the Department of Energy. R2304-DOE, June. RAND Corporation.

Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *American Economic Review* 101(5), 1649–75.

OECD Nuclear Energy Agency and International Atomic Energy Agency. 2009. *Uranium 2009: Resources, Production, and Demand.* ("The Red Book.") OECD, Nuclear Energy Agency.

Pindyck, Robert S. 1993. "Investment of Uncertain Cost." *Journal of Financial Economics* 34(1): 53–76.

Schmalensee, Richard. 1980. "Appropriate Government Policy toward Commercialization of New Energy Supply Technologies." *Energy Journal* 1(2): 1–40.

Thornton, Rebecca A., and Peter Thompson. 2001. "Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding." *American Economic Review* 91(5): 1350–68.

U.S. Atomic Energy Commission, Office of Planning and Analysis. 1974. *Nuclear Power Growth, 1974–2000.* WASH-1139(74). U.S. Department of Energy.

U.S. Department of Energy (DOE), Energy Information Administration. 1986. *An Analysis of Nuclear Power Plant Construction Costs.* DOE/EIA-0485. U.S. Department of Energy.

U.S. Department of Energy (DOE), Energy Information Administration. 1997. *Nuclear Power Generation and Fuel Cycle Report 1997.* DOE/EIA-0436(97). U.S. Department of Energy.

U.S. Department of Energy (DOE), Energy Information Administration. 2010. *Updated Capital Cost Estimates for Electricity Generation Plants.* November. U.S. Department of Energy. http://www.eia.gov/oiaf/beck_plantcosts/.

U.S. Department of Energy (DOE), Energy Information Administration. 2011a. "Annual Energy Review 2010." DOE/EIA-0384, October. U.S. Department of Energy.

U.S. Department of Energy (DOE), Energy Information Administration. 2011b. *Annual Energy Outlook 2011.* DOE/EIA-0383, April. U.S. Department of Energy.

U.S. Nuclear Regulatory Commission (NRC). 2011. *Information Digest 2011–2012.* NUREG-1350, Volume 23, August.

von Hippel, Frank N. 2011. "The Radiological and Psychological Consequences of the Fukushima Daiichi Accident." *Bulletin of the Atomic Scientists* 67(5): 27–36.

Wald, Matthew L. 2010. "Giant Holes in the Ground." *MIT Technology Review* November/December, http://www.technologyreview.com/energy/26542/.

Zimmerman, Martin B. 1982. "Learning Effects and the Commercialization of New Energy Technologies: The Case of Nuclear Power." *Bell Journal of Economics* 13(2): 297–310.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 67–92

The Private and Public Economics of Renewable Electricity Generation†

Severin Borenstein

The primary public policy argument for promoting electricity generation from solar, wind, and other renewable sources is the unpriced pollution externalities from burning fossil fuels. Some parties also advocate renewable electricity generation to improve energy security, price stability, or job creation, but these arguments are more difficult to support in a careful analysis, as I discuss later. Even comparing the higher costs of renewables with the environmental benefits, however, is not straightforward. Issues arise because the market value of electricity generation is very dependent on its timing, location, and other characteristics and because quantification of the nonmarket value from reduced emissions is difficult and controversial.

Since Pigou's (1920) seminal work, economists have understood that pricing externalities is likely to be the best way to move behavior towards efficiency. In the context of electricity, this insight means taxes on emissions or a tradable permit system, but such market-based policies have garnered limited political support in the United States and elsewhere. Instead, many governments have created policies to promote renewable electricity generation directly, through either subsidies or mandates. How well do these alternative policies substitute for pricing the negative externalities of generation from fossil fuel generation?

■ *Severin Borenstein is E. T. Grether Professor of Business Economics and Public Policy, Haas School of Business, University of California, Berkeley, California. He is a Co-Director of the Energy Institute at Haas, and Director of the University of California Energy Institute. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His e-mail address is ⟨borenste@haas.berkeley.edu⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.1.67.

doi=10.1257/jep.26.1.67

In this paper, I discuss the market and nonmarket valuation of electricity generation from renewable energy, as well as the costs and the subsidies that are available. On a direct cost basis, renewables are expensive, but the simple calculations don't account for many additional benefits and costs of renewables. I begin by briefly discussing studies of the costs of renewables and conventional generation, highlighting the primary cost drivers. I then discuss the many adjustments that are necessary to account for the time, location, and other characteristics that vary across and within generation technologies. Many such adjustments are idiosyncratic, differing substantially by individual project, but broader technology characteristics also play an important part in their determinations.

The next steps in the analysis, evaluating the benefits of reducing externalities with renewables, are more difficult than they may at first seem. The timing and location of renewable generation will affect what generation is displaced, as will the pre-existing mix of fossil fuel generation in the system (for a short-run analysis) or counterfactual mix (for a longer-run analysis). I then turn to other potential market failures that may affect the value that renewable energy offers and describe how these impact justifications for government policy such as job creation, industry building, energy security, and moderating swings in energy prices. I argue that these justifications are generally not supported empirically and in some cases are based on faulty economic reasoning.

In normative analyses of renewable electricity generation, there is often confusion about which economic actors are included in the welfare being evaluated. For instance, should a small town that is considering installing solar panels on city hall count federal subsidies as a benefit or just a transfer? Though economic analyses often draw a bright line between private and public benefits, renewable energy demonstrates that in practice there is a continuum of perspectives. Each may be appropriate for answering a different question. Evaluating the incentives of participants in a market generally requires doing the analysis from many perspectives.

Thus, I do not attempt here to rank order the benefit/cost ratios for the major technologies for generating electricity, which in any event will vary with the decision maker's preferences, the perceived costs of environmental externalities, and the state of technology. Technological progress, as well as ongoing research on externalities, make any such table obsolete shortly after it is printed. However, the microeconomic tools to carry out and critique such analyses are longer-lived. In this paper, I use the current issues in renewable energy cost analysis to illustrate the use, and occasional misuse, of those tools.

Generation Costs of Conventional and Renewable Energy

Though renewable sources of electricity generation other than hydroelectricity have grown very quickly in the last decade, they were starting from a miniscule base, and they remain a very small share of total generation today due primarily to their high direct cost. Table 1 presents the share of electricity generated from

Table 1
Electricity Generation by Source
(*units are billion kWh; data are for 2007)*

Region/Country	*Total*	*Natural gas*	*Coal*	*Nuclear*	*Hydro-electric*	*Oil and other liquids**	*Wind*	*Geo-thermal*	*Solar*	*Other renewables***
OECD										
OECD North America	**5,003**	**20%**	**44%**	**18%**	**13%**	**3%**	**0.8%**	**0.4%**	**0.0%**	**1.3%**
United States	4,139	22%	49%	19%	6%	2%	0.8%	0.4%	0.0%	1.3%
Mexico	244	37%	18%	4%	11%	26%	0.0%	2.9%	0.0%	1.2%
OECD Europe	**3,399**	**22%**	**29%**	**26%**	**15%**	**2%**	**2.9%**	**0.3%**	**0.1%**	**3.1%**
OECD Asia	**1,747**	**23%**	**40%**	**22%**	**7%**	**6%**	**0.3%**	**0.3%**	**0.0%**	**1.4%**
Japan	1,063	28%	31%	24%	7%	8%	0.2%	0.3%	0.0%	2.1%
Total OECD	*10,149*	*21%*	*38%*	*21%*	*12%*	*3%*	*1.4%*	*0.4%*	*0.1%*	*1.9%*
Non-OECD										
Non-OECD Europe and Eurasia	**1,592**	**36%**	**25%**	**17%**	**18%**	**4%**	**0.0%**	**0.0%**	**0.0%**	**0.1%**
Russia	959	40%	23%	15%	18%	3%	0.0%	0.0%	0.0%	0.2%
Non-OECD Asia	**4,779**	**10%**	**69%**	**2%**	**14%**	**4%**	**0.4%**	**0.3%**	**0.0%**	**0.3%**
China	3,041	2%	80%	2%	14%	2%	0.2%	0.0%	0.0%	0.1%
India	762	6%	71%	2%	16%	3%	1.4%	0.0%	0.0%	0.3%
Middle East	**674**	**57%**	**5%**	**0%**	**3%**	**35%**	**0.0%**	**0.0%**	**0.0%**	**0.0%**
Africa	**581**	**25%**	**45%**	**2%**	**17%**	**11%**	**0.2%**	**0.2%**	**0.0%**	**0.2%**
Central and South America	**1,009**	**15%**	**6%**	**2%**	**65%**	**9%**	**0.1%**	**0.3%**	**0.0%**	**2.6%**
Total Non-OECD	*8,634*	*20%*	*47%*	*5%*	*20%*	*7%*	*0.2%*	*0.2%*	*0.0%*	*0.5%*
Total world	**18,783**	**21%**	**42%**	**14%**	**16%**	**5%**	**0.9%**	**0.3%**	**0.0%**	**1.3%**

Source: EIA *International Energy Outlook 2010*, tables H11–H20.
* Includes petroleum-derived fuels and non-petroleum-derived liquid fuels, such as ethanol and biodiesel, coal-to-liquids, and gas-to-liquids. Petroleum coke, which is a solid, is included. Also included are natural gas liquids, crude oil consumed as a fuel, and liquid hydrogen.
** Includes biomass and other waste energy sources.

conventional and renewable sources for regions of the world and selected countries during 2007, the most recent year for which comparable worldwide data are available. Coal is the dominant generation source worldwide, with natural gas, hydroelectricity, and nuclear power also playing major roles.

Coal and natural gas remain the lowest-cost technology for new electricity generation in most parts of the world. These cost comparisons, however, show remarkable variance, with renewable generation far from competitive in some studies and quite cost-effective in others. Nearly all of these studies calculate a "levelized" cost of electricity, but as I discuss below, the exact economic assumptions made can drive enormous variation.

A User's Guide to Levelized Cost of Electricity Estimates

The levelized cost of electricity for a given generation plant is the constant (in real terms) price for power that would equate the net present value of revenue from the plant's output with the net present value of the cost of production.[1] Levelized cost estimates depend on numerous engineering factors that vary with the technology being reviewed, but these are not usually the main drivers of variation in estimates for a given plant. Current technological specifications for a plant are comparatively easy to establish with reasonable precision. For the most part, researchers agree on what inputs are going in and what outputs result.

Usually economic variables are behind the large discrepancies among levelized cost estimates. These include assumptions about inflation rates, real interest rates, how much the generator is going to be used, and future input costs, including fuel costs. Engineering factors also interact with these economic considerations; for example, the optimal usage of a plant will depend on the marginal cost of production, the speed with which its output can be adjusted, and the market price (plus other compensation, such as marginal subsidies) that the generator receives. The best levelized cost studies state these assumptions clearly, but many such studies do not.

Because generation plants are heterogeneous in location, architecture, and other factors, even plants with similar technology will not have the same levelized cost of electricity. The variation tends to be relatively small for coal and gas plants because the fuel is fairly standardized and the plant operation is less affected by location. Even the costs of these plants, however, are affected by idiosyncratic site characteristics (including property values), local labor costs, environmental constraints, access to fuel transportation, and access to electricity transmission lines, as well as variation in technical efficiency of operation. Production from solar and wind generation is largely driven by local climate conditions, and this greatly increases the variance in levelized cost across these types of projects.

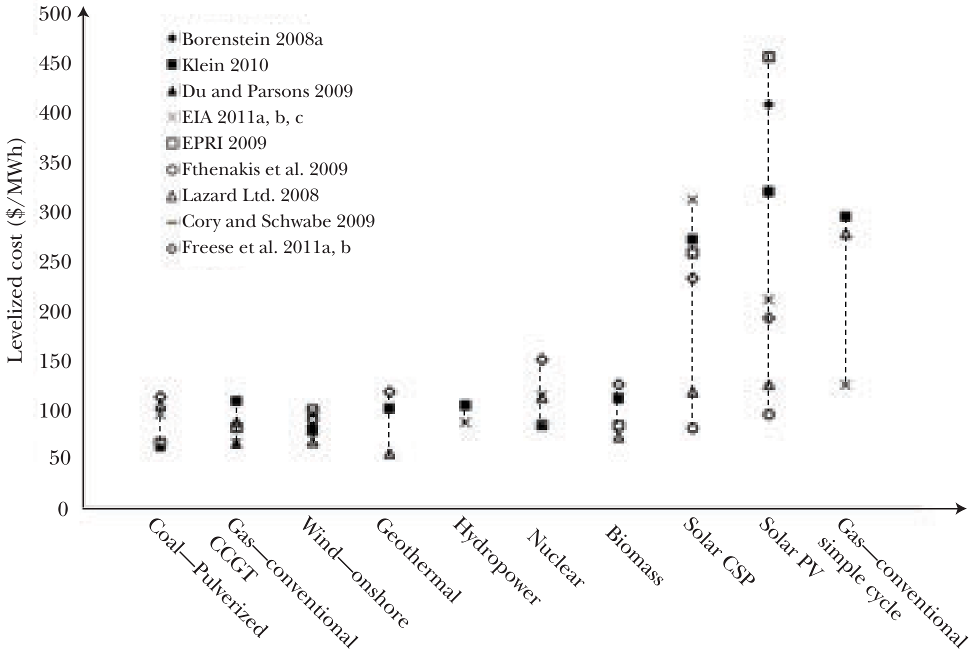
The variation in levelized cost across plants with the same technology raises an important caveat: levelized cost studies are usually based on the average outcome at existing or recent plants, but they are generally intended to guide future investment decisions. Technological progress, learning-by-doing, and economies of scale in building multiple plants will tend to make the cost of the marginal plant *lower* than the average of existing or recent facilities, while scarcity of high-quality locations will tend to make the cost of a new plant *higher* than the pre-existing average. Some studies are explicitly prospective, evaluating the levelized cost of a technology that the

[1] If a plant lasts N periods and produces q_n in period n, then discounting future cash flows at the real cost of capital r, the levelized cost of electricity is defined by

$$\sum_{n=1}^{N} q_n \frac{LCOE}{(1+r)^n} = \sum_{n=0}^{N} \frac{C_n(q_1,\ldots,q_N)}{(1+r)^n} \Leftrightarrow LCOE = \frac{\sum_{n=0}^{N} \frac{C_n(q_1,\ldots,q_N)}{(1+r)^n}}{\sum_{n=1}^{N} \frac{q_n}{(1+r)^n}}$$

where $C_n(q_1,\ldots,q_N)$ is the real (in period 0 dollars) expenditures in period n to produce the stream of output $(q_1,\ldots,q_N)$. As the formula suggests, this approach includes capital costs borne before any production can take place.

Figure 1
Levelized Cost Estimates



Notes: CCGT is "combined cycle gas turbine." CSP is "concentrated solar power." PV is "photovoltaic." MWh is "megawatt hour."

authors assume will be installed in some future year. These are necessarily the most speculative, forecasting future technological progress, which gives the authors great latitude to make varying assumptions that yield widely varying levelized cost estimates.

The lack of comparability in levelized cost analyses is particularly troubling because these cost figures are frequently the central focus of policy discussions about alternative technologies. These figures can potentially be useful benchmarks, but they must be thoughtfully adjusted for the attributes of the power produced and other impacts of the generation process.

I consider here only studies for U.S. generation. Costs vary around the world, both due to varying technologies and expertise, and because fuel costs and regulations differ.

Estimates of Levelized Costs of Electricity

With those cautions, Figure 1 presents levelized cost estimates for major electricity generation technologies. The notes in the Appendix at the end of this article present details of the calculations.[2] Clearly, the range of estimates can be significant

[2] Also see Intergovernmental Panel on Climate Change Working Group III (2011) for discussion of renewable energy technologies and Annex III for levelized cost estimates.

and the details in the Appendix table demonstrate why. Many of the studies include subsidies and tax benefits to the generator itself. With a sufficient subsidy, of course, any technology can appear to have a low cost. Nonetheless, these calculations can still be relevant for private decision making. A separate issue, which I discuss below, is accounting for upstream subsidies to fuel supply or transportation.

Coal and natural gas—the two leading sources of electricity generation—are fuel-intensive generation technologies (in terms of cost share) relative to the others, with natural gas being the most fuel-intensive major generation technology. (Oil-fired generation is even more fuel-intensive, but has a very small share of grid-connected generation in the United States due to its high cost.) Thus, forecasts of future fuel prices play a large role in levelized cost estimates. These forecasts have high variance due to uncertainty about the exhaustability of the resource, technological progress in exploration and extraction, and government regulation (Holland 2003).

Variation in technology and usage within generation technologies using the same fuel source can also greatly affect levelized cost. Combined-cycle gas turbine plants are highly efficient (in terms of "heat rate," the amount of fuel energy needed to generate a unit of electricity), but relatively costly to build, while single-cycle generation combustion turbine gas plants are less efficient but much cheaper to build. As a result, combined-cycle plants tend to run most of the time, while combustion turbines are used primarily at peak times, running far fewer hours per year. The levelized costs of these two technologies are quite different, but the comparison isn't informative, because they are intended for different uses. Because electricity demand is quite variable and electricity is not storable in a cost-effective way, there is demand for some "baseload" generation that runs in most hours and some "peaker" generation that is called on for relatively few hours per year. Neither technology could efficiently substitute for the other.

Hydroelectric and geothermal generation are generally viewed as renewable. They can be inexpensive, but locations that are usable and high productivity are quite limited. Large-scale hydroelectricity generation also creates such major alterations to the landscape that it is generally not considered environmentally friendly. In addition, hydroelectric generation usually faces a limit on total energy that can be produced in a year or other time frame due to precipitation and water storage limits.

The three broad categories of renewable energy that are considered closest to being scalable and cost competitive are wind, solar, and biomass. Wind and solar are also location-limited, though not to the same extent as hydro and geothermal. Studies have identified sufficient sites that if these locations were developed with wind and solar generation they could make these technologies the dominant electricity sources in the United States—see NREL (2010) on wind power and Fthenakis, Mason, and Zweibel (2009) on solar. The more significant barriers are cost of generation, cost of transmitting the power to where demand is, and the value of the power generated. The lowest-cost wind power is usually generated in fairly remote locations, so the cost of infrastructure to transmit the power to demand sites can be significant. Transmission costs for connection to the grid are generally not included in levelized cost estimates, in part because they are so idiosyncratic by project. Local resident resistance to transmission lines and incomplete property rights in some cases can also create significant regulatory uncertainty.

Solar power encompasses two different fundamental technologies. Solar thermal generation focuses sunlight on a heat transfer fluid that is used to create steam, which is then used in a turbine to drive a generator. Photovoltaic systems use semiconductors to convert sunlight directly to electricity.[3] Either technology can be used for large-scale generation in open space, known as utility-scale generation, while photovoltaic panels can be installed on a small-scale near demand, such as on residential rooftops.

Rooftop solar reduces the need for investment in high-voltage transmission lines that carry power from large-scale generation to local distribution wires. Some argue that it also reduces the cost of the local distribution networks, but there do not seem to be reliable studies on the distribution cost impact, as I discuss below. Economies of scale at the local distribution level are significant, suggesting the marginal savings from reduced flow on distribution lines is well below the average cost of distribution per kilowatt hour. Small-scale rooftop solar, such as on a single-family home, also enjoys fewer economies of scale in construction or panel procurement, so the up-front cost per unit of capacity tends to be much greater.[4]

Biomass is a broad category that includes both burning the inputs directly and biomass gasification, in which the inputs are heated to produce a synthetic gas. The primary biomass fuels are wood scraps and pulping waste, but also agricultural residue, landfill gas, and municipal solid waste. The levelized cost of biomass tends to depend to a great extent on the idiosyncratic local cost of collecting and preparing the fuel. In 2007, biomass provided about half of the non-hydro renewable electricity generation in the United States and the world. Mostly, this is from mixing biomass with coal and burning in a conventional coal-fired power plant, which requires fairly small incremental equipment investments. Such approaches represent the lower end of the levelized cost estimates in Figure 1, but the opportunity for expansion is limited.

Limitations of Using Levelized Cost Estimates to Compare Electricity Technologies

Although levelized cost in some form has been the starting point for cost comparisons since the beginning of electricity generation—McDonald (1962) discusses levelized cost comparisons from the early twentieth century—it is by no

[3] The International Energy Agency provides brief useful overviews of alternative energy technologies at ⟨http://www.iea.org/techno/essentials.htm⟩.

[4] To some extent, the lower panel cost for photovoltaic "farms" is a pecuniary economy, not representing real resource savings, if it is just a rent transfer from sellers to buyers. But to the extent that the panel cost is higher for small installations due to higher shipping or transaction costs of small orders, or because of the need to customize panel selection to particular types of installations, those probably reflect real cost differences.

means the final word. Difficulties arise because electricity generation technologies have different temporal and spatial production profiles.

Because electricity is very costly to store, wholesale prices can vary by a factor of 10 or more within a day. As a result, time variation in production, and the operator's control over that variation, greatly affects the value of power produced. Generation resources over which an operator has greater temporal control are considered "dispatchable," while those that vary significantly due to exogenous factors are considered "intermittent." Joskow (2011a, 2011b) discusses in detail the impact of temporal output variation on the value of power produced by different generating sources.

Among conventional gas and coal plants, there are constraints on how quickly a plant's output level can be increased or decreased ("ramping rates"), how long the plant must remain off once it has been shut down, and how frequently it must be shut down for planned or unplanned maintenance, and there is the cost of starting the plant. Economic tradeoffs also arise here between short-run benefits of pushing the plant to or beyond the engineering specifications and the longer-run costs of increased wear on the plant components that cause greater need for planned outages and greater incidence of unplanned outages.

Gas-fired peaker plants, for instance, have low fuel efficiency, but are very flexible, with rapid ramping capability and low start-up costs. Hydroelectric generation is also highly valued for its ability to adjust output very quickly. If the optimal "dispatch" of a plant implies that it will run disproportionately at times when electricity is of particularly high value—as is the case with gas-fired peaker generation and most hydro generation—then any levelized cost comparison must be augmented with adjustment for this enhanced value of the power that is produced.

Generation resources that depend on the local weather—such as wind and solar—are intermittent and therefore the least dispatchable. Such generation is almost entirely out of the control of the plant operator (although these technologies can be shut down fairly easily and quickly, so the plant operator can usually put an upper limit on their output). Power from intermittent resources must be evaluated in terms of the time at which it is produced. Solar power is produced only during daylight hours and tends to peak in the middle of the day. In many areas, this is close to coincident with the highest electricity demand, which usually occurs on summer afternoons. Thus, the average economic value of generation from solar is greater than if it produced the same quantity of power on average at all hours of the day. Wind power often has the opposite generation pattern in the United States, in most locations producing more power at night and at times of lower demand and prices.

Adjustment for the time variation of production is straightforward: compare the levelized cost to the average wholesale value of the power it delivers. In Borenstein (2008a), I find that power from solar photovoltaics in California is likely to be about 20 percent more valuable than the average power sold in the state, because it is produced disproportionately at high-priced times. The premium would be as high as 50 percent if the wholesale market were allowed to clear at very high prices, but that doesn't occur, because grid operators use "generation reserves," discussed below, to meet demand without allowing prices to rise too high at peak times. Fripp and Wiser (2008) find that wind power production in the West is likely to be 0–10 percent less valuable per unit than if the wind generator had the same average output in every hour of the day, though that study may understate the appropriate discount in wind value because it uses data from a period of very low power price volatility.

However, even this temporal adjustment for wholesale power prices doesn't completely capture the granularity over which the true value of power fluctuates. Because electricity is not storable at reasonable cost and the demand side of the market has had limited opportunity to respond to price fluctuations in very short time intervals, it is more cost-effective to build back-up generation in sufficient quantity to have most adjustment occur on the supply side of the market.[5] The presence of back-up generation in itself is not a barrier to efficient pricing that reflects the actual shadow value of power at each point in time, though the shadow value is likely to be low at most times. Grid operation, however, has never been based on such a precise market model. That approach made sense under the old utility model in which all generation was owned by one company, a company which solved a complex optimization problem and implemented the solution administratively. Even in the more than 20 years in which merchant (non-utility) generators have played a significant role in U.S. electricity markets, grid operators have generally just procured "reserve" generation services and charged it to the system as a whole. Thus, the cost to the system of an intermittent producer has been socialized across all generators and prices have not fully reflected the time-varying value of power. There is now an active debate about how much the failure to assign these costs of intermittency to specific generators skews incentives.

Adjusting levelized cost estimates for the intermittency depends on the degree to which intermittency requires additional reserve generation or increases the risk of a supply shortage that causes blackouts or brownouts. While a grid can easily handle very small shares of intermittent resources—in fact, to a grid operator they look almost the same as the stochastic component of demand that supply must follow—some grid engineers have argued that the cost will increase more than proportionally if intermittent resources constitute a significant share of generation, such as 20 percent or more, as is currently contemplated and has been achieved in some locations in Europe. This too is an area of active debate; a detailed discussion appears in New York Independent System Operator (2010). Ideal market pricing would reveal the value of a generator's production at every instant, but wholesale electricity markets are not set up to generate such fine-grained price signals.

[5] The technology for near-instantaneous demand response now exists, but questions remain about the cost-effectiveness of incorporating such sophisticated demand activity. If customers found it acceptable to have their thermostat respond automatically to retail price changes, that is, if they considered the associated cost to be fairly low, then the cost of intermittency could be substantially reduced. See Callaway (2009).

There is also a multiyear temporal issue that complicates comparisons of levelized costs. Levelized cost does not incorporate any variation in the real value of power across years. For instance, if the real cost of electricity is expected to rise substantially over time, then power produced in the near-term is less valuable than in the distant future. Comparing levelized costs implicitly assumes that the real marginal value of power will be constant. This assumption is particularly important if the output profiles of two generators differ substantially, such as comparing a nuclear plant that will take five to ten years to build to solar panels that will start producing within a year or less.

Just as the value of electricity varies temporally due to storage constraints, it also varies locationally due to transmission constraints. Complete locational pricing is difficult logistically due to the complex physics of power flows, but a number of areas of the United States do have what is known as "locational marginal pricing" that sends fairly efficient short-run price signals. The greater challenge in locational pricing is in the long run, because the full incremental cost of adding new transmission capacity can differ significantly from the direct infrastructure cost once one accounts for the resulting change in transmission capacity on all lines in the grid. Highly granular pricing—in both time and location—had less value in the historical electricity supply paradigm, with its lower reliance on intermittent generation and with its single utility that could coordinate long-term generation and transmission investment and internalize the externalities created by each in terms of grid capacity and intermittency. Even in the markets that remain regulated today, many of these issues still arise as regulated utilities buy much more power from independent generators than they did 10 or 20 years ago.

Locating electricity generation at the customer site, known as "distributed generation," engenders the most controversy in locational valuation. Retail prices are a very poor guide to locational value, because they include significant fixed cost recovery (for instance, the fixed costs of local distribution networks) and they reflect little or none of the locational (or time) variation in wholesale power purchase or production cost. At one extreme, some advocates of distributed solar and wind generation argue that customers should not only be able to reduce their power bills to zero by generating as much power over a billing period as they consume, they should be paid the retail rate by the utility for any net power they contribute to the system. At the other extreme, some grid engineers argue that intermittent distributed generation not only doesn't reduce local distribution costs much at all—so should be compensated no more than the wholesale price of power—the intermittent nature of the power and the reverse flow from customers increases the stress on distribution transformers and increases the frequency of repairs. At the heart of this conflict is an internal inconsistency in the utility revenue model: local electricity distribution service is a regulated, largely fixed-cost, business, but costs are recovered primarily through charges that vary with the quantity of electricity consumed. In the United States, wholesale electricity costs average only about 50 to 75 percent of residential retail electricity bills; most of the rest represents costs that don't vary with marginal electricity consumption.

Residential solar photovoltaic generation has been at the center of this debate. Residential solar does offer greater value than suggested by its high levelized cost—because it produces disproportionately at times of high demand, reduces transmission investment, and avoids the small percentage of power that is dissipated as heat when it is sent through the transmission and distribution lines from a distant generator (Borenstein 2008a). Nonetheless, *retail* rates don't accurately reflect the social value of distributed solar generation. With distributed generation, a significant share of the savings customers see in their electricity bills would have gone to pay the utility's fixed costs. These costs change very little, even in the long run, when customers generate some of their own power.

Adjustments for Subsidies and Preferential Tax Treatment

Some of the levelized cost estimates shown in Figure 1 (and described in detail in the Appendix) reflect costs after direct subsidies and preferential tax treatments, and some don't state clearly how subsidies and taxes are handled. Excluding subsidies and tax advantages seems sensible for cost analyses that are intended to guide public policy, but even that approach can be questioned. For instance, should state regulators consider federal subsidies and tax breaks when evaluating a proposed renewable energy facility? Given the political and logistical barriers to accomplishing Pareto-improving trades in these markets, the appropriate treatment will depend on whose welfare the decision maker weighs most heavily.

Excluding direct subsidies and tax breaks from levelized cost analyses is relatively straightforward, though it can be challenging in practice. Indirect subsidies that occur upstream and affect the price of inputs are more difficult to sort out. Advocates for renewable electricity argue that fossil fuel extraction receives special tax treatment in the United States. While that is likely true, and subsidies for fossil fuels are larger than for renewable energy in aggregate, the subsidy per kilowatt hour for fossil fuel generation is quite small. Adeyeye, Barrett, Diamond, Goldman, Pendergrass, and Schramm (2009) estimate that total subsidies for fossil fuels from 2002–2008 were \$72 billion in the United States, of which about \$21 billion plausibly went to domestically produced coal and natural gas that went into electricity production (explained in the online Appendix available with this paper at ⟨http://e-jep.org⟩). Even if these subsidies were passed through 100 percent to consumers, which seems highly unlikely for these internationally traded goods, that would amount to \$0.0011 per kilowatt hour of generation in the United States over this period. Other estimates of subsidies to coal and natural gas for electricity generation are substantially lower (EIA 2008) or many times higher (Koplow 2010), but over the range of subsidies claimed, the effect on electricity generation costs will not materially affect their comparison to renewable sources.

In 29 U.S. states and the District of Columbia, renewable energy benefits from a different sort of indirect subsidy, a minimum share of electricity that is mandated to come from renewables, often termed a "renewable portfolio standard." Nearly all such programs, however, translate this quantity standard to some extent into a subsidy/tax system through tradable credits for renewable energy, which can be purchased by retail electricity providers in lieu of meeting the standard through their own generation. As a result, some calculations of the economics of renewables may include the value of these credits. Whether such value should be counted in a social cost calculation depends on whether the credit price reflects the true cost of externalities avoided by the generation, which is difficult to assess, as I discuss in the next section. Schmalensee (forthcoming) discusses the different policies for promoting renewable energy generation and their effectiveness.[6]

With many factors affecting calculations of the full cost and benefit of generation technologies, claims that a new technology has attained "grid parity" must be interpreted with great caution. Advocates of wind generation who argue that it is at grid parity in some locations generally do not adjust for the timing, location, and intermittency factors that can make wind substantially less valuable. Residential solar photovoltaic power is sometimes claimed to be at grid parity if it saves the customer money (usually, after subsidies), but such analyses do not consider that the retail electricity rate pays for much more than just the energy that the solar generation replaces. Of course, grid parity on market factors alone is not the socially optimal driver of technology choice if some technologies produce greater negative externalities than others.

Incorporating Environmental Externalities

Until the 1960s, air pollution from conventional electricity generation was largely unregulated and in that sense "free" to the polluter. But in the 1960s and 1970s, legislation restricted the rights of generators to emit local air pollutants, particularly sulfur dioxide, nitrous oxides, and mercury. These policies didn't put prices on pollutants, but were command and control regulation, such as requiring the installation of smokestack devices ("scrubbers") that remove sulfur dioxide and other pollutants. In the last two decades, carbon dioxide has been found to be a major contributor to climate change, leading to efforts to restrict its emissions as well. About 33 percent of anthropogenic greenhouse gas emissions in the United States come from the electric power sector, with 27 percent coming from transportation, 20 percent from industry, and the remaining 20 percent is agriculture, commercial, or residential (EPA 2011, table 2-12).

In a first-best economic world, pollution rights would be just another input to the production of electricity from a given technology and would automatically be included in the levelized cost calculation. In most of the United States and the world, however, markets for rights to emit greenhouse gases or local pollutants are spotty at best. Most levelized cost estimates do not include the costs of emissions directly, though they do generally include the cost of technology that must be installed in order to meet command and control regulations.

[6] Also see ⟨http://www.dsireusa.org/⟩, a comprehensive database of such programs in the United States.

A large literature exists on the marginal social cost of the air pollutants that power plants emit. For local pollutants, the cost varies across plants and depends very much on the population density, climate, and geography around the plant, as well as the presence of other pollutants (Fowlie and Muller 2010). For greenhouse gases, the damage is not localized, so valuation is much more uniform across plants. All of these studies rely heavily on meteorological, climate, and public health models, as well as valuations of statistical lives. Muller and Mendelsohn (2007) explain the details and uncertainties of such studies and present estimates of the cost of local pollutants. The caveats applied to local pollution cost estimates are even stronger for estimates of the marginal social costs of greenhouse gas emissions because there is even more uncertainty in the underlying climate and public health models. Greenstone, Kopits, and Wolverton (2011) present a detailed discussion of the uncertainties in estimating the social cost of greenhouse gas emissions.

Absent government intervention, the external costs will not be borne by producers and will not affect choices among electricity generation technology. The obvious solution is to price the externalities—either through a tax or tradable permit program. The relative merits of these approaches have been debated at length (Keohane 2009; Metcalf 2009; and cites therein). Still, the reality is that both approaches remain relatively rare compared to alternative interventions such as technology mandates and subsidies for green power.

Technology mandates for pollution controls on conventional electricity generation have been and remain the most common response to these market failures. Technologies to remove some pollutants from the smokestack emissions of power plants have been used since the 1960s. It is well-known that such mandates can be inefficient, because they apply uniform standards to emitters with very different production profiles, costs of meeting the regulations, and costs of alternative technologies or production changes that would allow similar pollution reductions. Also known, but less highlighted, is that these command and control regulations don't account for whether the emissions occur at times when they are likely to be more or less damaging to public health. This is particularly important for nitrous oxides, which under some, but not all, meteorological conditions combine with volatile organic compounds and sunlight to make ozone. Even pricing the externality solves this problem only if prices reflect such variation, which is often not the case, generally for reasons of simplicity (Fowlie and Muller 2010).

Subsidies for green power (or mandated utility offer prices for power generated in this way, known as "feed-in tariffs") have been portrayed as nearly equivalent to pricing externalities, but more politically acceptable. This approach, however, is very problematic for three closely related reasons.

First, subsidizing green power for reducing pollution (relative to some counterfactual) is not equivalent to taxing "brown" power to reflect the marginal social damage. If end-use electricity demand were completely inelastic and green and brown power were each completely homogeneous, they would have the same effect; the only effect of the subsidy would be to shift the production *share* towards green and away from brown power. But the underlying market failure is the underpricing of brown power, not the overpricing of green power, so subsidizing green power from government revenues artificially depresses the price of power and discourages efficient energy consumption.[7] As a result, government subsidies of green power lead to overconsumption of electricity and disincentives for energy efficiency. In addition, for any given level of reduction, it will be achieved more efficiently by equalizing the marginal price of the pollutant across sectors as well as within sectors. This is not achievable through *ad hoc* subsidies to activities that displace certain sources of emissions. Fowlie, Knittel, and Wolfram (forthcoming) estimate that failure to achieve uniform marginal prices in the emissions of nitrogen oxides in the United States has raised the cost of regulation by at least 6 percent.

Second, subsidizing green power generally fails to recognize the heterogeneity within the green power sector and among the brown power sources that are being displaced. Solar power that reduces coal-fired generation lowers greenhouse gas emissions by about twice as much on average as if it reduces natural-gas-fired generation. Assuming that the marginal generation displaced is equal to the average generation mix in the system can be a poor approximation. A number of studies have attempted to go further and infer the generation that is displaced by an incremental unit of power from wind or solar within a system, accounting for the timing and location of the green power (for example, Callaway and Fowlie 2009; Cullen 2011; Gowrisankaran, Reynolds, and Samano 2011). These studies have made it clear how difficult it is to identify the alternative generation emissions even after the fact. But to give efficient long-run incentives for investment, policymakers must commit to subsidies well before they could have the data to calculate the alternative emissions. The problem arises because subsidizing green power is an indirect approach to the pollution problem, and the relationship between green power and emissions avoided is not uniform. It would not arise with a direct tax (or pricing through tradable permits) on pollution.[8]

Third, because subsidizing green power addresses the policy goal only indirectly, it introduces an opportunity for what might be called "benefit leakage" in which the effect on the policy goal takes place out of the immediate area. If producing more green power in one state lowers the production of brown power in a distant area that exports electricity to the state, then the benefits of the pollution reduction are less likely to flow to those underwriting the subsidies. Obviously, with greenhouse gases this would be an accounting issue, not a real change in the benefits, but with local pollutants the local environmental gains from subsidizing green power could be much less than would be suggested by a calculation that assumes no change in trade.

[7] Green power subsidies that are paid for through a general surcharge on electricity are likely to be a step in the right direction, but only in very special cases do they result in electricity prices that reflect the social cost of pollution.

[8] Both subsidizing green and taxing brown power require committing to the level of a policy instrument—such as prices or quantities—with only imperfect knowledge of its optimal level. Subsidizing green has the *additional* problem of setting the level of the policy instrument while knowing only imperfectly the relationship between the policy instrument and the variables of real interest.

All energy sources have environmental implications for which property rights have not been clearly assigned or would be costly to enforce. Wind turbines harm birds, as well as create low-frequency thumping that some people find difficult to live with. Large-scale solar projects in the desert can endanger habitat for native animals. Solar photovoltaic panels contain some heavy metals that require careful handling in disposal. Geothermal generation may cause ground water pollution and small-scale seismic activity. Tidal and wave power—both in nascent development stages—will likely run into concern that the generators interfere with marine life. Coal mining creates significant quantities of solid waste. Oil and gas production can result in leaks that spoil nearby ecosystems. Recently there have been concerns about the environmental impact of fluids used in hydraulic fracturing. Nearly all generation sources are at some point accused of visual pollution.

Many of these externalities involve substantial costs which mean substantial wealth transfers and potentially large efficiency implications. Externalities from fossil fuels have triggered litigation for years. With each new energy source, new property rights conflicts emerge and must be adjudicated. Even if Coasian efficiency results after property rights are assigned, the assignment process is costly. In one vivid example in Sunnyvale, California, a conflict arose between one neighbor with solar panels and another with redwood trees that had grown tall enough to shade the panels. After a lengthy lawsuit, the solar panels won and the redwood trees had to be removed (Rogers 2008).

Non-Environmental Externalities

While environmental externalities are the leading argument for public policy that encourages alternative energy sources, they are certainly not the only argument made. Although these non-environmental justifications have become more prominent in public policy discussions in the last year or so, they are generally much less persuasive.

Energy Security

"Energy security" is rarely defined precisely, but the phrase generally is used to suggest that the United States should produce a higher share of the energy it uses. One justification is macroeconomic: If the price of a fuel for which the United States is highly import-dependent rises suddenly, the common wealth shock to most consumers could potentially disrupt the macroeconomy. Empirically, this argument may apply to oil—the United States now consumes nearly twice as much oil as it produces—but it does not apply to coal or natural gas, for which the United States is about self-sufficient. Moreover, the United States uses almost no oil in producing electricity. Energy security arguments could perhaps support a move towards electric cars (or other alternatives to oil for transportation fuels). In that case, however, producing the electricity from coal or natural gas enhances security as much as producing it from renewables. In addition, electricity from coal and natural gas is less expensive, so using those sources would make electric transportation more affordable. The distinct advantage of renewable electricity generation is its lower environmental impact, not its ability to enhance energy security.

A second "energy security" argument is that high energy prices enrich some energy-exporting countries that are hostile to U.S. global interests. By reducing use of these fuels, the argument goes, the United States could lower the price of energy, which would both help United States consumers and reduce the wealth flows to hostile regimes and possibly reduce military expenditures directed towards ensuring unimpeded energy trade. This argument again does not have traction in analysis of coal or natural gas in the United States. Even in oil-importing countries where oil is a significant source of electricity generation, the quantities of oil used for generation are so small relative to the world oil market that replacing them with renewables is unlikely to have any noticeable impact on world oil prices, as indicated in Table 1. This argument has been raised with more credibility in the context of European natural gas purchases from Russia.

Non-Appropriable Intellectual Property

Even with the strong intellectual property laws that have been adopted in the most advanced countries, in most cases a successful innovator captures relatively little of the value from the innovation. That outcome surely creates some dynamic inefficiency, which governments have addressed in many sectors by subsidizing basic research. Whether this incentive problem is greater in energy than other sectors is not clear, but it is clear that U.S. government expenditures on energy R&D have been much smaller as a share of GDP contribution than in health care, defense, or technology (NSF 2010 Chapter 4).

Government support for generating fundamental scientific knowledge in energy has increased with the creation of the Advanced Research Projects Agency—Energy (ARPA-E) within the Department of Energy in 2009, but the ARPA-E budget for 2012 is likely to be under $200 million. Studies from across the political spectrum have suggested it should be 50 percent to many times higher (Augustine et al. 2011; Loris 2011).

For those renewable electricity technologies currently available, a common argument for subsidies is that greater installation will lead to learning-by-doing that will drive down the cost and price of the technology. This possibility justifies government intervention, however, only if the knowledge from that learning-by-doing is not appropriable by the company that creates it—that is, if the knowledge spills over to other firms. Though the argument has some merit, proponents frequently overstate the strength of the evidence on this point.

First, most studies of learning-by-doing are not able to separate learning-by-doing from other changes. In solar photovoltaic power, costs have come down dramatically since the 1960s as the total number of installed panels has increased, with estimates that every doubling of the installed base has on average been associated with about a 20 percent decline in the cost of solar panels (for instance, Duke and Kammen 1999; Swanson 2006). Many factors have affected costs over this time

(Nemet 2006). Significant exogenous technological advances in crystalline silicon solar technologies have resulted from investments made outside the commercial solar power sector, especially public investments made as part of the U.S. space program and private investments in the semiconductor industry. In addition, firms in the industry have gotten larger, which has lead to savings from economies of scale—producing more units of output in each period—rather than learning-by-doing, which is the knowledge gained from a larger aggregate history of production over time. Barbose, Dargouth, Wiser, and Seel (2011) present comprehensive data on changing costs of solar photovoltaics over time.

The distinction between learning-by-doing and economies of scale may seem minor, but the implications for public policy are immense. If one firm can drive down its costs by producing at large scale in its factory or its installation operation, those benefits are highly appropriable by that large firm. Smaller firms are not likely to experience a cost decline because a competitor is enjoying economies of scale. Thus, significant economies of scale in any industry, short of creating a natural monopoly, are not generally seen as a basis for government intervention.

Learning-by-doing creates more spillovers, because knowledge is likely to be portable across firms. Still, the evidence of strong learning-by-doing is thin and credible results on spillovers are even more rare. Nemet's (2006) analysis suggests that learning-by-doing has actually played a relatively small role in the decline of solar photovoltaic costs over the last 30 years. He finds that the scope for learning-by-doing using the current crystalline silicon technology is quite limited given the current state of the industry. While the evidence of minimal learning-by-doing effects in solar photovoltaics is not dispositive, it is more convincing than any existing research claiming significant effects.

Green Jobs

The "job creation" justification for government policies to promote renewable energy took on greater prominence after the downturn that began in 2007 and the failure of climate change legislation in Congress since then. In the green jobs debate of 2008–2010, there was much confusion between the short-run stimulus goal and the longer-run policy of subsidizing green job creation. As a stimulus program, the advisability of subsidizing renewable energy depends on how rapidly the investment can take place and the elasticity of investment with respect to those subsidies. In general, the renewable energy sector tends to require large up-front construction costs, which is likely to be attractive in the context of short-term job creation, but the capacity to expand such projects rapidly is likely to be fairly limited.

When the economy recovers and the stimulus justification fades, is there a longer-term job creation justification for subsidizing renewable energy? This question has a static and a dynamic component. The static view is that renewable energy and energy efficiency are more labor-intensive technologies for producing (or conserving) energy than conventional energy production. The empirical support for these claims is uneven, but even if true, it is far from making the case that green job creation is welfare improving. To the extent that renewable energy costs more, even after accounting for environmental externalities, renewable energy absorbs more resources to produce the same value of output—a unit of electricity—and lowers GDP compared to conventional sources. Another possibility is that renewable energy creates "better" jobs than conventional sources, perhaps by targeting workers whose incremental economic welfare is of particular importance because they are otherwise difficult to employ or because they would otherwise have very low-wage jobs.

The dynamic view is that investment in renewable energy is justifiable as an attempt to change the equilibrium path of investment and the economy. One reason suggested is that renewable energy is a growth industry and, implicitly, that private investors are too slow to recognize the opportunity, leading to suboptimal investment. However, it seems hard to argue the general case that government policymakers are better at identifying emerging business opportunities than the private sector. A more nuanced and potentially compelling version of this argument is that up-front investment will create network externalities and learning that spill over much more strongly intra-nationally than internationally, creating a sustainable economic advantage for the country that makes the investment (Moretti 2012). Such effects could be important, but as countries make competing investments to become the dominant center of renewable energy, it seems likely that at least some of those rents would be dissipated or transferred to firms that can choose their locations.

The network effects argument is often heard in political debates, but evidence supporting it is scarce. Both Germany and Spain have subsidized enormous investments in installation of renewable energy, particularly solar. In 2008, Spain was the largest market for new solar generation in the world, but its manufacturing and installation of new capacity virtually disappeared in 2009 when the country cut back subsidies. Germany has continued to grow installations of solar photovoltaics, more than quadrupling new capacity from 2008 to 2010, but panel manufacturing in Germany has declined from 77 percent of new installed capacity in 2008 to 27 percent in 2010 as China and Taiwan have made massive investments in panel manufacturing, according to data from Earth Policy Institute (2011).

This area is ripe for further research. I am not aware of any credible studies that have assessed the short-run stimulus effect of green energy investment relative to other stimulus policies, the quality of the jobs created in the long run by green energy investment, or the ability of governments to make strategic investments that trigger a sustainable new sector.

Lowering the Cost of Fossil Fuel Energy

Increasing adoption of renewable energy lowers the demand for fossil fuels and drives down their prices. As a public policy argument, this is essentially advocating the exercise of monopsony power in the fossil fuel market (a buyer or set of buyers can drive down price by reducing purchases). That outcome has clear inefficiencies—some fossil fuels are replaced by more-expensive renewable power—but it still might be surplus-enhancing on net for the set of economic actors that the policymaker represents. In the United States, the effect of increasing renewable power is

Table 2
Levelized Cost of Residential Solar Photovoltaic Power under Alternative Discount Rates
(per kilowatt hour)

Real interest rate	*1%*	*3%*	*5%*	*7%*	*9%*
Levelized cost	$0.249	$0.315	$0.389	$0.468	$0.551

Notes: Table 2 presents the implied levelized cost of power for a 5 kilowatt system located in Sacramento, California. Assumptions: A five kilowatt system costs $36,500 installed. Panels last for 30 years with no shading or soiling and no maintenance costs, producing on average 0.77 kilowatts over all hours in first year. Output of panels declines by 0.5 percent per year due to aging. The inverter is replaced after 10 years (at $2,552) and 20 years (at $2,171), based on current cost of $3000 declining by 2 percent annually in real terms. For further details and sources, see online Appendix.

to reduce demand for natural gas and coal. U.S. production of these fossil fuels is nearly equal to consumption, so the effect is to transfer wealth from U.S. producers to U.S. consumers. On the state level within the United States, the effect is much more uneven since many states are large importers of fossil fuels and a smaller number are large exporters.

The size of this effect on prices is also questionable. While some advocates have focused on short-run price variation, the impact of a long-term shift towards renewables will depend on the long-run elasticity of supply for natural gas and coal. With the advent of hydraulic fracturing, it seems likely that the long-run elasticity of natural gas supply has become quite high. The long-run elasticity of coal supply is generally seen as quite high as well (Miller, Wolak, and Zhang 2011). Thus, a shift to renewables is not likely have a large effect on fossil fuel prices.

An Application to Residential Solar Photovoltaic Power

Here, I apply the analytic approach described above to update the calculations of levelized cost of residential solar power from Borenstein (2008a), taking into account recent changes in the cost of solar photovoltaic systems.

According to Barbose, Dargouth, Wiser, and Seel (2011), the cost of installing residential-scale solar systems (less than 10 kilowatt capacity) in 2010 varied in average price from $6.3/watt in capacity in New Hampshire to $8.4/watt of capacity in Utah, with California—by far the largest state for residential solar—at $7.30. Taking California's number as the benchmark, Table 2 presents the implied levelized cost of power for a 5 kilowatt system located in Sacramento, California, under alternative real discount rates. The underlying assumptions noted in the table are intended to be median estimates: if anything, they are tilted somewhat towards lowering the cost.

The real interest rate of 3 percent implies a levelized cost of $0.315 per kilowatt hour. I follow Borenstein (2008a) in adjusting for the timing of production (including line losses), increasing the value of residential solar by 20 percent, and for the location of production, increasing value by 1 percent. I incorporate these effects by adjusting the levelized cost down to \$0.260 per kilowatt hour $(= 0.315/(1.2 \times 1.01))$. An additional downward adjustment of \$0.02 per kilowatt hour accounts for long-run savings in transmission investment, as discussed in Borenstein (2008b, p. 10), which brings the net cost to \$0.240. Details of these adjustments are in the online Appendix available with this paper at ⟨http://e-jep.org⟩. This result compares to levelized costs for combined-cycle gas-fired generation that are now generally below \$0.08 per kilowatt hour, given the reduced price forecasts for natural gas that are now common due to the expected supply increases.

Adjusting next for environmental externalities, if one assumes that new residential solar generation substitutes for new combined-cycle gas turbines, then the local pollutant reduction is valued at about \$0.0015 per kilowatt hour according to Muller, Mendelsohn, and Nordhaus (2011). That leaves a cost gap between residential solar and combined-cycle gas turbine generation of at least \$0.158. The gas plant emits slightly less than 0.0005 tons of carbon dioxide per kilowatt hour of electricity, so residential solar would be cost competitive on a social cost basis only if the cost of carbon dioxide emissions were greater than \$316 per ton. Nearly all social cost and price forecasts for carbon dioxide are well below \$100 per ton (Greenstone, Kopits, and Wolverton 2011), which leaves residential solar still at least \$0.108 per kilowatt hour more expensive.

This analysis of the costs of residential solar power does not account for potential cost savings in reducing the size of the necessary distribution network for electricity, nor for spillovers from learning-by-doing, for which analyses offer much less guidance. On the other side, it also doesn't incorporate reduced output due to shading or soiling of the panels, or installation at a less-than-ideal angle due to the building orientation (Borenstein 2008b). But this analysis provides a good notion of the gap that those factors would have to fill in order for residential solar photovoltaics to substitute cost-effectively for gas-fired generation.

Medium-scale and large-scale solar photovoltaics installations and large-scale solar thermal generation are somewhat more cost competitive. Contracts for these larger systems are not public, but reports in the industry press suggest the unsubsidized levelized cost from these installations is probably between \$0.15 and \$0.20 per kilowatt hour in 2011, before any of the market or externality adjustments and likely using more than a 3 percent real cost of capital. These systems enjoy the same production timing benefit as residential solar, but less (or none) of the reduction in line losses and transmission savings. These systems would require a much lower cost of carbon dioxide to be competitive with gas-fired generation, though still probably \$100 per ton or greater.

Conclusion

The most important market failure in energy markets is almost certainly environmental externalities, and the single most efficient policy would be to price those externalities appropriately. Yet policymakers often find pricing externalities to be nearly impossible politically. Thus, the second-best discussion is over which, if any, alternative policy interventions are likely to do the most good, or at least to do more good than harm.

Instead of pricing externalities, the far more prevalent government response has been targeted programs to promote specific alternatives to conventional electricity generation technologies. Justifications for such programs have generally begun with environmental concerns, but have often expanded to energy security, job creation, and driving down fossil fuel prices, generally without support of sound economic analysis. Such targeted programs also seem especially vulnerable to political manipulation.

If governments are to implement reasoned renewable generation policy, it will be critical to understand the costs and benefits of these technologies in the context of modern electricity systems. This requires developing sophisticated levelized cost estimates, and adjusting for both the market value of the power generated and the associated externalities, so they can be usefully compared across projects and technologies. Such adjustments are complex and frequently controversial. More research at the interface of the economics and engineering of electricity markets would be very valuable, particularly on the cost of intermittency, the benefits of end-use distributed generation, and the economic spillovers from learning-by-doing and network externalities. Progress on these questions would enhance renewable energy public policy and private decision making, particularly in a world where first-best, market-based options are greatly restricted.

■ *I am grateful to Judd Boomhower for excellent research assistance. I also benefited from the comments of Duncan Callaway, Lucas Davis, Meredith Fowlie, Michael Greenstone, Bill Hogan, Paul Joskow, Chris Knittel, Karen Notsund, Richard Schmalensee, and the editors. This research was supported in part under a research contract from the California Energy Commission to the Energy Institute at Haas.*

References

Adeyeye, Adenike, James Barrett, Jordan Diamond, Lisa Goldman, John Pendergrass, and Daniel Schramm. 2009. *Estimating U.S. Government Subsidies to Energy Sources: 2002–2008.* Washington, DC: Environmental Law Institute.

Augustine, Norm, Ursula Burns, John Doerr, Bill Gates, Charles O. Holliday, Jeff Immelt, and Tim Solso (American Energy Innovation Council). 2011. *Catalyzing American Ingenuity: The Role of Government in Energy Innovation.* Washington, DC: American Energy Innovation Council.

Barbose, Galen, Naïm Dargouth, Ryan Wiser, and Joachim Seel. 2011. "Tracking the Sun IV: An Historical Summary of the Installed Cost of

Photovoltaics in the United States from 1998 to 2010." Lawrence Berkeley National Laboratory Paper LBNL-5047E. http://eetd.lbl.gov/ea/emp/reports/lbnl-5047e.pdf.

Borenstein, Severin. 2008a. "The Market Value and Cost of Solar Photovoltaic Electricity Production." Center for the Study of Energy Markets Working Paper 176. http://www.ucei.berkeley.edu/PDF/csemwp176.pdf.

Borenstein, Severin. 2008b. "Response to Critiques of 'The Market Value and Cost of Solar Photovoltaic Electricity Production'." http://faculty.haas.berkeley.edu/borenste/SolarResponse.pdf.

California Energy Commission. 2007a. "2007 Final Natural Gas Market Assessment." CEC-200-2007-009-SF. In Support of the 2007 Integrated Energy Policy Report. Final Staff Report. http://www.energy.ca.gov/2007publications/CEC-200-2007-009/CEC-200-2007-009-SF.pdf.

California Energy Commission. 2007b. *2007 Integrated Energy Policy Report.* CEC-100-2007-008-CMF. California Energy Commission. http://www.energy.ca.gov/2007publications/CEC-100-2007-008/CEC-100-2007-008-CMF.pdf.

Callaway, Duncan S. 2009. "Tapping the Energy Storage Potential in Electric Loads to Deliver Load Following and Regulation, with Application to Wind Energy." *Energy Conversion and Management* 50(5): 1389–1400.

Callaway, Duncan, and Meredith Fowlie. 2009. "Greenhouse Gas Emissions Reductions from Wind Energy: Location, Location, Location?" http://nature.berkeley.edu/~fowlie/papers.html.

Cory, Karlynn, and Paul Schwabe. 2009. "Wind Levelized Cost of Energy: A Comparison of Technical and Financing Input Variables." National Renewable Energy Laboratory Technical Report NREL/TP-6A2-46671. http://www.nrel.gov/docs/fy10osti/46671.pdf.

Cullen, Joseph A. 2011. "Measuring the Environmental Benefits of Wind-Generated Electricity." http://www.u.arizona.edu/~jcullen/Documents/measuringwind.pdf.

Davis, Lucas, and Catherine Wolfram. 2011. "Deregulation, Consolidation, and Efficiency: Evidence from U.S. Nuclear Power." EI @ Haas Working Paper 217, Energy Institute at Haas.

Du, Yangbo, and John E. Parsons. 2009. "Update on the Cost of Nuclear Power." MIT Center for Energy and Environmental Policy Research Paper 09-004. http://web.mit.edu/mitei/docs/spotlights/nuclear-fuel-cycle-du.pdf.

Duke, Richard, and Daniel M. Kammen. 1999. "The Economics of Energy Market Transformation Programs." *Energy Journal* 20(4): 15–64.

Earth Policy Institute. 2011. "Annual Installed Solar Photovoltaics Capacity in Selected Countries and the World, 1998–2010" and "Annual Solar Photovoltaics Production by Country, 1995-2010." Excel Worksheets available at http://www.earth-policy.org/data_center/C23.

Electric Power Research Institute (EPRI). 2009. "Program on Technology Innovation: Integrated Generation Technology Options." http://my.epri.com/portal/server.pt?Abstract_id=000000000001019539.

Energy Information Administration (EIA). 2008. *Federal Financial Interventions and Subsidies in Energy Markets 2007.* http://www.eia.gov/oiaf/servicerpt/subsidy2/pdf/subsidy08.pdf.

Energy Information Administration (EIA). 2010. *International Energy Outlook 2010.* U.S. Energy Information Administration. http://www.eia.gov/oiaf/ieo/.

Energy Information Administration (EIA). 2011a. *Annual Energy Outlook 2011.* U.S. Energy Information Administration. http://www.eia.gov/forecasts/aeo/.

Energy Information Administration (EIA). 2011b. *Assumptions to the Annual Energy Outlook 2011.* U.S. Energy Information Administration. http://205.254.135.24/forecasts/aeo/assumptions/.

Energy Information Administration (EIA). 2011c. "Levelized Costs in the Annual Energy Outlook 2011." November. http://205.254.135.24/oiaf/aeo/electricity_generation.html.

Environmental Protection Agency (EPA). 2011. *Inventory of U.S. Greenhouse Gas Emissions and Sinks: 1990–2009.* Washington, DC: Environmental Protection Agency.

Fowlie, Meredith, Christopher R. Knittel, and Catherine Wolfram. Forthcoming. "Sacred Cars? Optimal Regulation of Stationary and Nonstationary Pollution Sources." *American Economic Journal: Economic Policy.*

Fowlie, Meredith, and Nicholas Muller. 2010. "Designing Markets for Pollution when Damages Vary across Sources: Evidence from the NO_x Budget Program." http://nature.berkeley.edu/~fowlie/fowlie_muller_ASSA.pdf.

Freese, Barbara, Steve Clemmer, Claudio Martinez, and Alan Nogee. 2011a. *A Risky Proposition: The Financial Hazards of New Investments in Coal Plants.* Union of Concerned Scientists. http://www.ucsusa.org/assets/documents/clean_energy/a-risky-proposition_report.pdf.

Freese, Barbara, Steve Clemmer, Claudio Martinez, and Alan Nogee. 2011b. Appendix A "Key Assumptions for Levelized Cost of Electricity Ranges" to *A Risky Proposition: The Financial Hazards of New Investments in Coal Plants.* http://www.ucsusa.org/assets/documents/clean_energy/Appendix-Key-Assumptions-Levelized-Costs.pdf.

Fripp, Matthias, and Ryan H. Wiser. 2008. "Effects of Temporal Wind Patterns on the Value of Wind-Generated Electricity in California and the Northwest." *IEEE Transactions on Power Systems* 23(2): 477–85.

Fthenakis, Vasilis, James Mason, and Ken Zweibel. 2009. "The Technical, Geographical, and Economic Feasibility for Solar Energy to Supply the Energy Needs of the US." *Energy Policy* 37(2): 387–99.

Gowrisankaran, Gautam, Stanley S. Reynolds, and Mario Samano. 2011. "Intermittency and the Value of Renewable Energy." NBER Working Paper 17086.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. 2011. "Estimating the Social Cost of Carbon for Use in Federal Rulemakings: A Summary and Interpretation." NBER Working Paper 16913, March.

Holland, Stephen. 2003. "Set-up Costs and the Existence of Competitive Equilibria when Extraction Capacity Is Limited." *Journal of Environmental Economics and Management* 46(3): 539–56.

Intergovernmental Panel on Climate Change Working Group III. 2011. *Special Report on Renewable Energy Sources and Climate Change Mitigation*. http://srren.ipcc-wg3.de/. IPCC.

Joskow, Paul L. 2011a. "Comparing the Costs of Intermittent and Dispatchable Electricity Generating Technologies." *American Economic Review* 101(3): 238–41.

Joskow, Paul. 2011b. "Comparing the Costs of Intermittent and Dispatchable Electricity Generation Technologies." Center for Energy and Environmental Research Working Paper 2010-013, revised February 2011.

Keohane, Nathaniel O. 2009. "Cap and Trade, Rehabilitated: Using Tradable Permits to Control U.S. Greenhouse Gases." *Review of Environmental Economics and Policy* 3(1): 42–62.

Klein, Joel. 2010. *Comparative Costs of California Central Station Electricity Generation Technologies*. California Energy Commission CEC-200-2009-017-SD. http://www.energy.ca.gov/2009publications/CEC-200-2009-017/CEC-200-2009-017-SF.pdf.

Koplow, Doug. 2010. "EIA Energy Subsidy Estimates: A Review of Assumptions and Omissions." Earth Track, Inc. http://earthtrack.net/files/uploaded_files/EIA%20subsidy%20review%20final_17Mar10.pdf.

Lazard Ltd. 2008. "Lazard Cost of Energy Analysis: Version 2.0." http://www.narucmeetings.org/Presentations/2008%20EMP%20Levelized%20Cost%20of%20Energy%20-%20Master%20June%202008%20(2).pdf.

Loris, Nicolas D. 2011. "Department of Energy Spending Cuts: A Guide to Trimming President Obama's 2012 Budget Request." Heritage Foundation Backgrounder No. 2545, April.

Margolis, Robert M. 2003. "Photovoltaic Technology Experience Curves and Markets." Paper presented at the NCPV and Solar Program Review Meeting, Denver, CO.

McDonald, Forrest. 1962. *Insull*. Chicago: University of Chicago Press.

Metcalf, Gilbert E. 2009. "Designing a Carbon Tax to Reduce U.S. Greenhouse Gas Emissions." *Review of Environmental Economics and Policy* 3(1): 63–83.

Miller, Michael, Frank Wolak, and Zhe Zhang. 2011. "Quantifying the Economic and Environmental Impacts of China's Buying Power in the Global Steam Coal Market." Program on Energy and Sustainable Development Working Paper, Stanford University, September.

Moretti, Enrico. 2012. ***The New Geography of Jobs.*** New York: Houghton Mifflin Harcourt.

Muller, Nicholas Z., and Robert Mendelsohn. 2007. "Measuring the Damages of Air Pollution in the United States." *Journal of Environmental Economics and Management* 54(1): 1–14.

Muller, Nicholas Z., Robert Mendelsohn, and William Nordhaus. 2011. "Environmental Accounting for Pollution in the United States Economy." *American Economic Review* 101(4): 1649–75.

National Renewable Energy Laboratory (NREL). 2010. "Windpowering America: Estimates of Windy Land Area and Wind Energy Potential, by State, for Areas >= 30% Capacity Factor at 80m." An Excel file: http://www.windpoweringamerica.gov/docs/wind_potential_80m_30percent.xlsx.

Nemet, Gregory F. 2006. "Beyond the Learning Curve: Factors Influencing Cost Reductions in Photovoltaics." *Energy Policy* 34(17): 3218–32.

New York Independent System Operator. 2010. *Growing Wind: Final Report of the NYISO 2010 Wind Generation Study*.

Pigou, Arthur C. 1920. *The Economics of Welfare*. London: Macmillan and Co.

Rogers, Paul. 2008. "Local Tree Law Balances Solar vs. Shade, Lawmakers Respond to Celebrated Sunnyvale Feud." *San Jose Mercury News,* July 23.

Schmalensee, Richard. Forthcoming. "Evaluating Policies to Increase the Generation of Electricity from Renewable Energy." *Review of Environmental Economics and Policy*.

Swanson, Richard M. 2006. "A Vision for Crystalline Silicon Photovoltaics." *Progress in Photovoltaics: Research and Applications* 14(5): 443–53.

Appendix

Table A1
Details for Levelized Cost of Energy Estimates

Borenstein 2008a	**Inflation:** All calculations use 2007$; **Interest**: 3% real annual interest rate; **Lifetime**: 25 years; **Capacity factor**: 16%; **Subsidies:** None; **Online:** 2007; **Notes:** Capacity factor is for AC production, based on production simulation for Sacramento, CA. Levelized cost of energy (LCOE) in real 2007$.
Klein 2010	**Inflation:** About 1.6% per year, plus 0.5% escalation for O&M costs; **Interest:** 4.67% weighted average cost of capital (WACC) for publicly-owned utilities; **Carbon cost:** None; **Local pollutant cost:** None; **Online:** 2018; **Notes:** LCOE given is in nominal terms. Used "average" case. Used publicly-owned utility estimates. ***Gas CCGT:*** **Lifetime**: 20 years; **Fuel**: $6.56/MMBtu in 2009 to $16.80/MMBtu in 2029, at nominal prices; **Capacity factor:** 75%. ***Wind – onshore:*** **Lifetime:** 30 years; **Capacity factor:** 37%; **Subsidies:** Federal production incentive of $4.10/MWh. ***Geothermal:*** **Lifetime:** 30 years; **Capacity factor:** 94%; **Subsidies:** Federal production incentive of $4.10/MWh. ***Hydropower:*** **Lifetime:** 30 years; **Capacity factor:** 30%; **Subsidies**: None; **Online:** 2018; **Notes:** For "small-scale and existing sites." ***Biomass:*** **Lifetime:** 20 years; **Fuel**: $2.00/MMBtu in 2009 to $2.91/MMBtu in 2029, at nominal prices; **Capacity factor:** 85%; **Subsidies:** Federal production incentive of $4.10/MWh; **Online:** 2018; **Note:** Data are for stoker boiler. ***Solar CSP & Solar PV:*** **Lifetime:** 20 years; **Capacity factor:** 27%; **Subsidies:** Receives federal production incentive of $4.10/MWh, and exempt from state ad valorem tax; **Notes:** 250 MW gross capacity parabolic trough for solar CSP (concentrated solar power); 250 MW gross capacity single axis system for PV (photovoltaic). ***Gas – conventional simple cycle:*** **Lifetime:** 20 years; **Fuel:** $6.56/MMBtu in 2009 to $16.80/MMBtu in 2029, at nominal prices; **Capacity factor:** 75%.
Du and Parsons 2009	**Inflation:** 3% annual inflation, plus 1% real escalation in O&M and 0.5% real escalation in fuel; **Lifetime:** 40 years; **Capacity factor:** 85%; **Subsidies:** None; **Online:** 2009; **Notes**: Real LCOE in 2007$. ***Pulverized:*** **Interest:** 7.8% real WACC; **Fuel:** $2.60/MMBtu in 2007$ with escalation as described above; **Notes:** Based on recently proposed supercritical and ultrasupercritical pulverized coal plants. ***Gas – Conventional CCGT***: **Interest:** 7.8% real WACC; **Fuel:** $7/MMBtu in 2007$ with escalation as described above. ***Nuclear***: **Interest:** 10% real WACC; **Fuel:** $0.67/MMBtu in 2007$ with escalation as described above.

(*continued on next page*)

EIA 2011a, b, c	**Inflation:** Average 2.9% annually; **Interest:** 10.4% real WACC for fossil generators without CCS, 7.4% real WACC for all others; **Lifetime:** 30 years; **Subsidies:** None; **Carbon cost:** Cost of capital for fossil plants without CCS is 3 percentage points higher than for other generators; **Online:** 2016; **Notes:** LCOE is in 2009$. ***Pulverized:*** **Fuel:** Delivered price is about $2.50/MMBtu in 2009$ through 2035; **Local pollutant cost:** Plants choose least-cost combination of scrubbers and emissions allowances to comply with Clean Air Interstate Rule; **Capacity factor:** 85%. ***Gas – Conventional CCGT*:** **Fuel:** Lower 48 wellhead price rises from about $4/kCF in 1990 to about $6.50/kCF in 2035; **Capacity factor:** 87%. ***Wind – onshore:*** **Capacity factor:** 34%. ***Geothermal:*** **Capacity factor:** 92%. ***Hydropower:*** **Capacity factor:** 52%. ***Nuclear:*** **Fuel:** Proprietary model starting from Energy Resources International uranium price forecasts; **Capacity factor:** 90%. ***Biomass:*** **Fuel:** Not given; **Capacity factor:** 83%. ***Solar CSP:*** **Capacity factor:** 25%. ***Solar PV:*** **Capacity factor:** 25%; **Notes:** For 150 MW fixed-tilt flat plat PV. ***Gas – conventional simple cycle:*** **Fuel:** Lower 48 wellhead price rises from about $4/kCF in 1990 to about $6.50/kCF in 2035; **Capacity factor:** 30%.
EPRI 2009	**Inflation:** All calculations use real 2008$; no escalation is modeled for any cost component; **Interest:** Real, after-tax WACC of 5.5%; **Lifetime:** 30 years; **Subsidies:** None; **Carbon cost:** none; **Online:** 2015; **Notes:** LCOE in 2008$. ***Pulverized:*** **Fuel:** $15/MWh in 2008$; **Capacity factor:** 80%; **Local pollutant cost:** Mercury removal; **Notes:** For 650–750 MW supercritical plant. ***Gas – Conventional CCGT:*** **Fuel:** $8–$10/MMBtu in 2008$; **Capacity factor:** 80%. ***Wind – onshore:*** **Capacity factor:** 35%; **Notes:** 100 MW wind farm; location not specified. ***Nuclear:*** **Fuel:** $0.80/MMBtu in 2008$; **Capacity factor:** 90%; **Notes:** 1400 MW plant. ***Biomass:*** **Fuel:** $1.22–$2.22/MMBtu in 2008$; **Capacity factor:** 85%; **Notes:** 75 MW circulating fluidized bed plant, with 28% efficiency. ***Solar CSP:*** **Capacity factor:** 32%; **Notes:** 125 MW facility in New Mexcio with wet cooling and 10% combustion. ***Solar PV:*** **Capacity factor:** 26%; **Notes:** 20 MW fixed flat plate PV with 10% conversion efficiency.
Fthenakis, Mason, and Zweibel 2009	**Inflation:** 1.9% annual; **Interest:** 6.7% after-tax WACC; 5% real discount rate; **Lifetime:** 30 years; **Subsidies:** not specified; **Online:** 2020; **Notes:** Assumes new HVDC transmission construction costs of $0.007/kWh. ***Solar CSP:*** **Capacity factor:** 90% (16 hours of thermal storage); **Notes:** “Gigawatt scale” CSP plant in southwest US with 16 hours of thermal storage capacity. ***Solar PV:*** **Capacity factor:** 90% (300 hours of compressed air storage); **Notes:** “Multi-hundred MW scale” PV; assumes major technological advances lower cost.

(*continued on next page*)

Lazard Ltd. 2008	**Inflation:** 2.5% annual escalation for fuel, O&M, and tax credits (no overall inflation specified); **Interest:** 7.3% after-tax WACC; **Lifetime:** 20 years; **Notes:** Online years imputed based on stated construction times; LCOE in 2008$; **Carbon cost:** None; **Local pollutant cost:** None. ***Pulverized coal:*** **Fuel:** $2.50/MMBtu in 2008$, with escalation as described above; **Capacity factor:** 85%; **Online:** 2013; **Notes:** Range of estimates $74–$135/MWh (high end includes 90% carbon capture and compression). ***Gas – Conventional CCGT:*** **Fuel:** $8.00 / MMBtu in 2008$; **Capacity factor:** 40%–85%; **Online:** 2011; **Notes:** Range $73–$100/MWh. ***Wind – onshore:*** **Capacity factor:** 28%–36%; **Subsidies:** Production tax credit of $20/MWh; **Online:** 2009; **Notes:** 100 MW facility; Range $44–$91/MWh. ***Geothermal:*** **Capacity factor:** 70%–80%; **Subsidies:** Production tax credit of $20/MWh; **Online:** 2011; **Notes:** Range $42–$69/MWh. ***Nuclear:*** **Fuel:** $0.50/MMBtu in 2008$; **Capacity factor:** 90%; **Online:** 2014; **Range:** $98–$126. ***Biomass:*** **Fuel:** $0–2/MMBtu in 2008$; **Capacity factor:** 80%; **Subsidies:** Production tax credit of $10/MWh; **Online:** 2012; **Notes:** Range $50–$94/MWh. ***Solar CSP:*** **Capacity factor**: 26%–38%; **Subsidies:** 30% investment tax credit; **Online**: 2010; **Notes:** Range $90–$145/MWh (low end tower, high end trough). ***Solar PV:*** **Capacity factor:** 20%–26%; **Subsidies:** 30% investment tax credit; **Online:** 2009; **Notes:** Range $96–$154/MWh; low end is for 10 MW net capacity thin film installation; high end is for 10 MW crystalline fixed axis installation. ***Gas – Conventional simple cycle:*** **Fuel:** $8.00/MMBtu in 2008$ with escalation as described above; **Capacity factor:** 10%; **Online:** 2010; **Notes:** Range $221–$334; Low end is for GE 7FA turbine; High end is for GE LM6000PC turbine.
Cory and Schwabe 2009	**Inflation:** 4% annually; **Interest:** All equity financing with 10% target internal rate of return; Interest rate 5.8%; **Lifetime:** 20 years; **Capacity factor:** 34%; **Subsidies:** Production tax credit of $15–$21/MWh; **Online:** 2008; **Notes:** 120 MW facility; Used "corporate" financing structure and "base case" scenario.
Freese, Clemmer, Martinez, and Nogee, 2011a, b	**Inflation:** All calculations use real 2010$. **Interest:** Not specified. Capital costs based on EIA technology-specific fixed charge rates; **Lifetime:** 20 years; **Subsidies:** None; **Carbon cost:** None; **Online:** 2015; **Notes:** LCOE is given in 2010$. ***Pulverized coal:*** **Fuel:** $1.60–$2.70/MMBtu in 2010$; **Capacity factor:** 85%; **Notes:** Range of estimates $103–$130/MWh; for 600 MW supercritical plant. ***Gas – conventional combined cycle:*** **Fuel:** $4.00–$6.75/MMBtu in 2010$; **Capacity factor:** 50%–87%; **Notes:** Range $52–$98/MWh; 400 MW plant. ***Wind – onshore:*** **Capacity factor:** 25%–45%; **Notes:** Range $57–$125/MWh. ***Geothermal:*** **Capacity factor:** 85%; **Notes:** Range $65–$169/MWh. ***Nuclear:*** **Fuel:** $0.8/MMBtu in 2010$; **Capacity factor:** 80%–90%; **Notes:** Range $141–$184/MWh; for 1100–1350 MW plant. ***Biomass:*** **Fuel:** $1.88–$4.06/MMBtu in 2010$; **Capacity factor:** 80%; **Notes:** Range $147–$328/MWh; 50 MW circulating fluidized bed plant. ***Solar CSP:*** **Capacity factor:** 27%–43%; **Notes:** Range $147–$328/MWh; 50–100 MW facility. ***Solar PV:*** **Capacity factor:** 20%–28%; **Notes:** Range $126–$260/MWh.

Notes: LCOE is "levelized cost of energy." WACC is "weighted average cost of capital." O&M is "operation and maintenance." CCS is "carbon capture and storage." CSP is "concentrated solar power." PV is "photovoltaic." CCGT is "combined cycle gas turbine." HVDC is "high-voltage, direct current."

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 93–118

Reducing Petroleum Consumption from Transportation

Christopher R. Knittel

The United States consumes more petroleum-based liquid fuel per capita than any other OECD high-income country—30 percent more than the second-highest country (Canada) and 40 percent more than the third-highest (Luxembourg). The transportation sector accounts for 70 percent of U.S. oil consumption and 30 percent of U.S. greenhouse gas emissions. Gasoline and diesel fuels alone account for 60 percent of oil consumption. The economic argument for seeking to reduce this level of consumption of petroleum-based liquid fuel begins with the externalities associated with high levels of U.S. consumption of petroleum-based fuels.

First, burning petroleum contributes to local pollution. The transportation sector accounts for 67 percent of carbon monoxide emissions, 45 percent of nitrogen oxide (NO_X) emissions, and 8 percent of particulate matter emissions. These pollutants lead to health problems ranging from respiratory problems to cardiac arrest. Furthermore, automobiles emit both NO_X and volatile organic compounds which, combined with heat and sunlight, form ground-level ozone, or smog. The papers by Currie and Walker (2011) and Knittel, Miller, and Sanders (2011) both find that decreases in traffic reduce infant mortality.

Second, burning a gallon of gasoline causes roughly 25 pounds of carbon dioxide to be emitted into the atmosphere, which raises the risks of destructive climate change. Greenstone, Kopits, and Wolverton (2011) estimate the social cost

■ *Christopher R. Knittel is the William Barton Rogers Professor of Energy Economics, Sloan School of Management, and a Research Associate, Center for Energy and Environmental Policy Research, Massachusetts Institute of Technology, and a Research Associate, National Bureau of Economic Research, all in Cambridge, Massachusetts. His e-mail address is ⟨knittel@mit.edu⟩.*

doi=10.1257/jep.26.1.93

of carbon under a variety of assumptions. They estimate a social cost of carbon as high as $65 per metric ton of carbon dioxide (and gases with an effect equivalent to carbon dioxide) in 2010, though their values are often in the range of $21 to $35 per metric ton. In Tol's (2008) metastudy of the existing literature since 2001, he finds the median social cost of carbon ranges from $17 to $62 per metric ton of CO_2 equivalent.

The U.S. dependence on imported gasoline has had other costs, too, including the military expense of trying to assure stability in oil-producing regions (for example, ICTA 2005), and the relationship between oil price shocks and macroeconomic downturns. These, too, can be viewed as negative externalities. But this paper neither focuses on these various externalities and social costs, nor delves into the literature about quantifying them. Instead, I take their existence as largely given and focus on understanding the policy tools that seek to reduce gasoline consumption.

Of course, an obvious starting point for economists is to look at prices: although the price of petroleum is set in a global market, government taxes on petroleum vary quite substantially. Table 1 lists taxes on gasoline and diesel on a per gallon basis as of 2010 for "OECD Category I countries"—essentially the world's most developed economies. The United States and Canada are clearly outliers, with taxes on gasoline below $1 per gallon. How do these price differences affect consumption? Figure 1 is suggestive. For each of these countries, it plots the per capita petroleum-based liquid fuel consumption versus the gasoline price in the country, with the size of the bubbles proportional to population. The regression line is population weighted, but looks similar if it is not weighted. It would require quite a bit of additional argument and delicacy to estimate a reliable elasticity of demand from these data, but for the record, the slope of a fitted log-log regression line through these data is –1.86. If one were to also include the log of income as an explanatory variable in such a regression, the coefficient associated with the log of gasoline prices is –1.49, while the coefficient associated with the log of income is 1.05.

The relative fuel use across the United States and other OECD Category I countries is, at least in part, a by-product of differences in the types and use of light-duty vehicles. Schipper (2006) reports that the average gallons-per-mile of European fleets in 2005 was below 0.034 (29.4 miles per gallon), while the average gallons-per-mile of the U.S. fleet was 0.051 (19.6 miles per gallon). Because of differences in how fuel efficiency is evaluated, this finding probably understates the European advantage. Similarly, Schipper reports that per capita miles traveled in European countries is between 35 to 45 percent of U.S. miles traveled.

The next four sections of this paper examine the main channels through which reductions in U.S. oil consumption might take place: 1) increased fuel economy of existing vehicles, 2) increased use of non-petroleum-based, low-carbon fuels, 3) alternatives to the internal combustion engine, and 4) reduced vehicle miles traveled. I then discuss how these policies for reducing petroleum consumption compare with the standard economics prescription for using a Pigouvian tax to deal with externalities. Taking into account that energy taxes are a political hot button in the United States, and also considering some evidence that consumers may not

Table 1
Motor Fuel Taxes for OECD Category I Countries in 2010
($/gallon)

Country	*Gasoline*	*Diesel*
United States	$0.49	$0.59
Canada	$0.96	$0.77
New Zealand	$1.20	$0.00
Australia	$1.34	$1.34
Iceland	$2.28	$2.03
Japan	$2.59	$1.55
Korea	$2.64	$1.87
Spain	$2.66	$2.08
Hungary	$2.68	$2.17
Austria	$2.77	$2.18
Luxembourg	$2.90	$1.94
Czech Republic	$3.04	$2.59
Switzerland	$3.09	$3.15
Slovak Republic	$3.23	$2.31
Sweden	$3.24	$2.56
Ireland	$3.41	$2.82
Italy	$3.54	$2.65
Belgium	$3.58	$2.10
Denmark	$3.58	$2.68
Portugal	$3.65	$2.28
France	$3.80	$2.69
Greece	$3.82	$2.39
Norway	$3.87	$2.97
Finland	$3.93	$2.28
United Kingdom	$3.95	$3.95
Germany	$4.10	$2.95
Netherlands	$4.19	$2.29

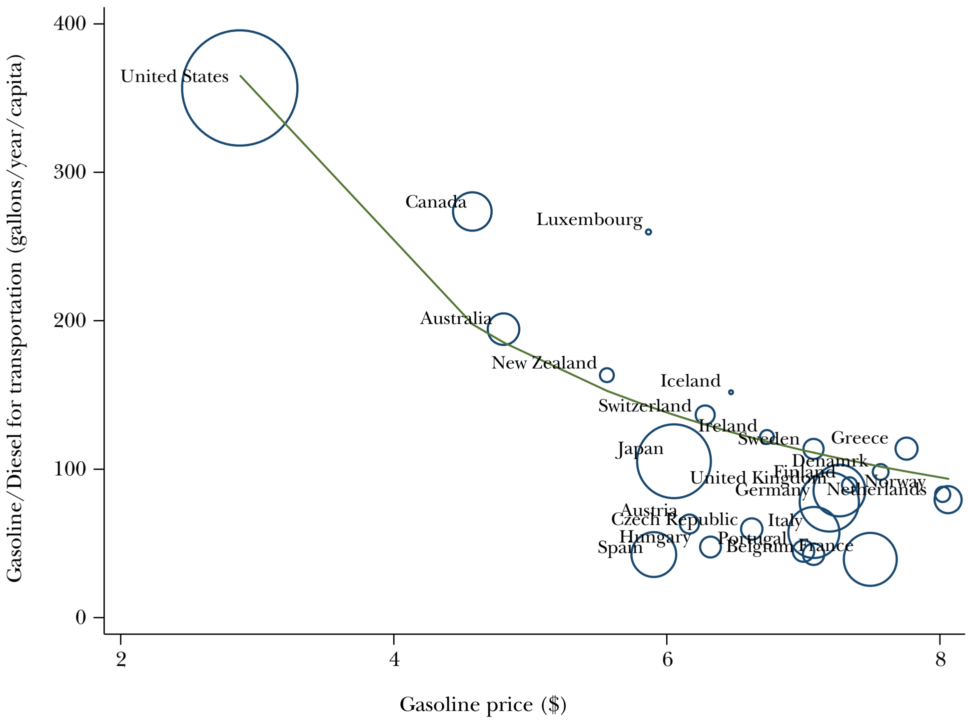
Source: Taken from an Alternative Fuels and Advanced Vehicles Data Center (AFDC) worksheet: ⟨www.afdc.energy.gov/afdc/data/docs/fuel_taxes_by_country.xls⟩. AFDC's source is the OECD/EEA database on instruments for environmental policy: ⟨http://www2.oecd.org/ecoinst/queries/index.htm⟩.
Notes: Rates as of January 1, 2010. Data for the United States and Canada include average excise taxes at the state/provincial level. VAT is not included.

"correctly" value fuel economy, I offer some thoughts about the margins on which policy aimed at reducing petroleum consumption might usefully proceed.

Improved Fuel Economy

Shortly after the oil price shocks of the 1970s, the United States adopted Corporate Average Fuel Economy (CAFE) standards, which set minimum average fuel economy thresholds for the new vehicles sold by an automaker in a given

Figure 1
Transportation Fuel Consumption per Capita versus Fuel Price

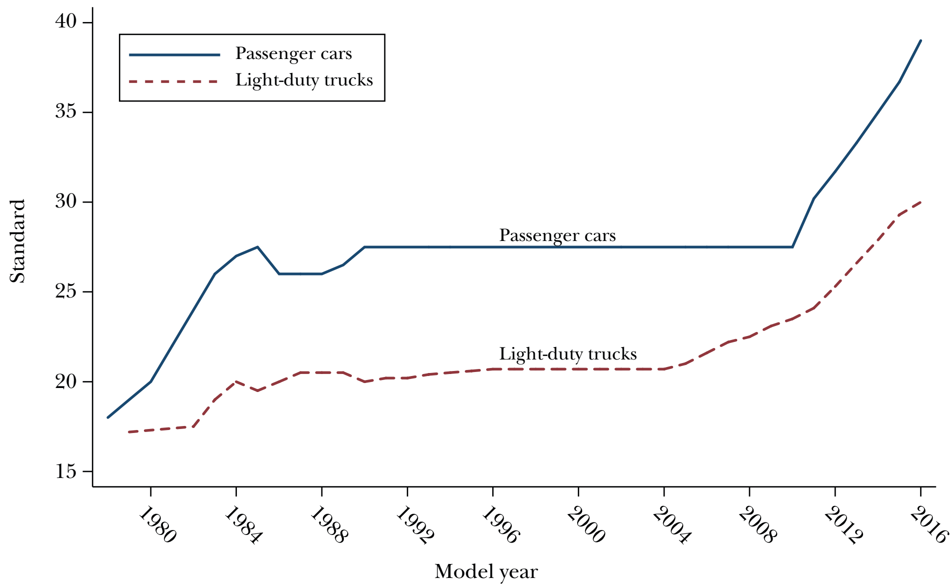


Source: Data from Worldbank.org.
Notes: Size of the circle proportional to population. The line is the fitted value from a regression of the log of consumption on the log of price.

year. Figure 2 shows how the standard evolved. For passenger cars, the standard increased by only 0.5 miles per gallon (MPG) from 1984 to 2010; for light-duty trucks, the increase was only 3.5 MPG over this same time period. From 1978 to 1991 the standard for light trucks differentiated between two- and four-wheel drive trucks, but manufacturers could also choose to meet a combined-truck standard. By world standards, these miles-per-gallon standards are not aggressive. After accounting for differences in the testing procedures, the World Bank estimated that the European Union standard was roughly 17 MPG more stringent in 2010 than the U.S. standard (An, Earley, and Green-Weiskel 2011).

Manufacturers who violate the CAFE standard pay a fine of roughly $50 per mile-per-gallon per vehicle. Historically, U.S. manufacturers have complied with the standard. Asian manufacturers have typically exceeded the standard in each year, while European manufacturers have typically violated the CAFE standard and paid the fines. Trading between manufacturers was not allowed, so there was no possibility for certain manufacturers to accumulate credits for

Figure 2
U.S. CAFE Standards from 1978 to 2016



Source: Data are from the National Highway Traffic Safety Administration.

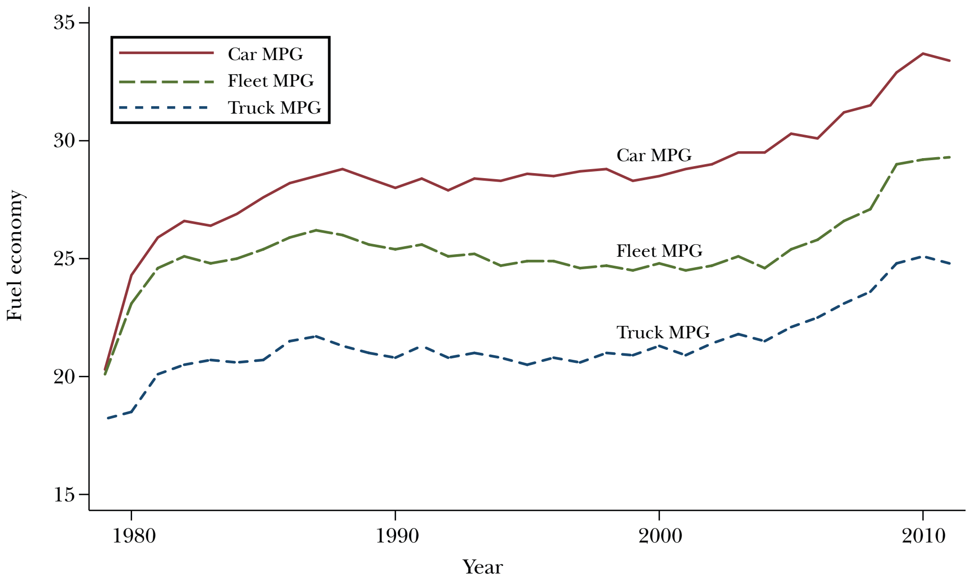
selling a higher proportion of fuel-efficient cars and then selling those credits to other manufacturers.

Other than the fact that the standards have barely budged over the last three decades, two features of the original CAFE standards reduced their effect. First, sport-utility vehicles were treated as light trucks, and thus could meet a lower miles-per-gallon standard than cars. Perhaps not coincidentally, in 1979 light trucks comprised less than 10 percent of the new vehicle fleet, but this share rose steadily and peaked in 2004 at 60 percent. Second, vehicles with a gross vehicle weight of over 8,500 pounds, which includes many large pickup trucks and sports-utility vehicles, were exempt from CAFE standards.

Actual new vehicle fleet fuel economy in the United States has changed little since the early 1980s. Figure 3 plots the fuel economy of passenger vehicles (cars) and light duty trucks from 1979 to 2011. The figure shows that while the average fuel economy of both cars and trucks increased over this time period, fleet fuel economy fell as consumers shifted away from cars and into trucks. The figure also shows that during the run-up in gasoline prices beginning in 2005, fleet fuel economy increased. This rise appears to have subsided by 2010.

Although the fuel economy of new U.S. vehicles gradually declined through the late 1980s and the 1990s, there was scope for substantial improvements. In the short run, when the set of offered vehicles is fixed, car buyers could choose vehicles with higher fuel efficiency. In 2011, for example, while the mean passenger car

Figure 3
U.S. New Vehicle Fuel Economy from 1979 to 2011



Source: Data are from the National Highway Traffic Safety Administration.

available for sale was rated at 23 miles per gallon, 10 percent of passenger cars had a rating of 30 MPG or above. The highest rating for 2011 was the Nissan Leaf at 99 MPG; the Toyota Prius had a combined fuel economy rating of 50 MPG.

In the medium run, automakers can adjust vehicle attributes by trading off weight and horsepower for increased fuel economy. In Knittel (2011), I find that reducing weight by 1 percent increases fuel economy by roughly 0.4 percent, while reducing horsepower and torque by 1 percent increases fuel economy by roughly 0.3 percent.

In the long run, manufacturers can push out the frontier. In Knittel (2011), I estimate that had manufactures put all of the technological progress observed in the market from 1980 to 2006 into fuel economy, instead of putting it into attributes that increased horsepower and/or weight, average fuel economy would have increased by 60 percent, instead of the 11.6 percent increase actually observed. On average, a vehicle with a given weight and engine power level has a fuel economy that is 1.75 percent higher than a vehicle with the same weight and horsepower level from the previous year. While the analysis in Knittel (2011) ends in 2006, using similar data and empirical models through model year 2011, the technological frontier has shifted out at an average rate of 1.97 and 1.51 percent per year from 2006 to 2011 for passenger cars and light-duty trucks, respectively, suggesting that no technological barrier has yet been reached. The greater availability of hybrids and plug-in hybrids also suggests that progress is likely to continue.

Gasoline prices do seem to affect choices about which cars to buy. A number of papers have used this variation in gasoline prices to estimate the magnitude of this response. These papers inevitably estimate a short-run response to gasoline prices—in the particular sense that the choice set of vehicles is usually held fixed. In Busse, Knittel, and Zettelmeyer (2011), my coauthors and I estimate that over the period 1999–2008, the market share of vehicles in the bottom quartile of fuel efficiency, among those vehicles offered in a given year, falls by nearly 24 percent for every $1 increase in gasoline prices. In contrast, the market share of the upper quartile of vehicles as ranked by fuel efficiency increases by over 20 percent. We also show that the market share of compact cars increases by 24 percent for every $1 increase in gasoline prices, while the market share of sport-utility vehicles falls by 14 percent. Klier and Linn (2010) estimate a logit demand system and focus on the effects of changes in a vehicle's cost per mile on demand. They find that a 5 cent increase in a vehicle's cost per mile, equivalent to a $1 increase in gasoline prices for a 20 miles-per-gallon vehicle, decreases the log of its market share by between 0.5 and 0.8, all else equal.[1] In the aggregate, this translates into an increase in average fuel economy of between 0.5 and 1.2 miles per gallon for every $1 dollar increase in gas prices. Again, their estimates hold the set of offered vehicles fixed. Li, Timmins, and van Haefan (2009) find similar effects.

A new CAFE standard in place for 2011 seeks to increase average fuel economy to roughly 34.1 miles per gallon by 2016. The Environmental Protection Agency and Department of Transportation are currently in the rule-making process for model years 2017 and beyond, with President Obama and 13 automakers agreeing to a standard of 54.5 MPG by 2025. A number of notable changes have occurred. First, the mileage standards are now based to some extent on the greenhouse gas emissions of the vehicle, which can deviate from fuel economy because of ancillary greenhouse gas emissions associated with, for example, air conditioner refrigerant leaks. Second, the new standards are "footprint"-based, in which each vehicle faces a standard based on the area of the footprint of its tires; larger footprints face a lower standard. For example, the 2011 Honda Civic coupe has a footprint of 43 square feet, while the 2011 Ford F-150 SuperCab has a footprint of 67 square feet. In 2016, these vehicles would face fuel efficiency standards of 41.1 MPG and 24.7 MPG, respectively.[2] For more details on the fuel efficiency rules for the next few years, see U.S. Energy Information Administration (2007) and U.S. Environmental Protection Agency (2010).

Is new vehicle fuel economy of 34.1 and 54.5 miles per gallon in 2016 and 2025, respectively, attainable? If we take the average rates of technological progress from Knittel (2011) and a new vehicle fuel economy in 2010 of roughly 29 MPG, new

[1] These coefficients represent the change in a vehicle's log of market share when the vehicle's cost per mile increases. Of course, if this is driven from changes in gasoline prices, then the cost of *all* vehicles' cost per mile will change. This explains why these effects are so large.

[2] See ⟨http://www.nhtsa.gov/cars/rules/cafe/overview.htm⟩. The sticker fuel economy is roughly 80 percent of how the vehicle is counted for the CAFE standard.

vehicle fuel economy in 2016 would be roughly 32 MPG in 2016, close to the standard of 34.1 MPG. Using the estimated trade-off coefficients, getting to 34.1 MPG would require reducing weight and engine power by less than 6 percent. Alternatively, increasing the rate of technological progress to 2.75 percent per year would achieve the mark.

And, what about the standard of 54.5 miles per gallon in 2025? Taken literally, it *would* require fundamental changes to rates of technological progress and/or the size and power of vehicles. The 2025 number is a bit misleading. In the law, the 54.5 miles-per-gallon standard is based on a calculation from the Environmental Protection Agency based on carbon dioxide tailpipe emissions. It also includes credits for many technologies including plug-in hybrids, electric and hydrogen vehicles, improved air conditioning efficiency, and others. On an apples-to-apples basis, Roland (2011) cites some industry followers that claim that the actual new fleet fuel economy standard in 2025 is more like 40 miles per gallon. Achieving 40 miles per gallon by 2025 is certainly possible. At a rate of technological progress of 1.75 percent per year, 40 miles per gallon requires additional reductions in weight and engine power of less than 7 percent.

Alternative Fuels

Biofuels are derived from biological components like corn, soybeans, sugar, grasses, and wood chips. The ethanol produced by this process is an imperfect substitute for gasoline, although biodiesel is a nearly perfect substitute for petroleum-based diesel. (Methanol is another alcohol and imperfect substitute for gasoline that can be derived from either methane—that is, natural gas—biomass, or coal.) Biofuels also hold the potential to have lower carbon emissions. If the plant material could be grown and converted to liquid fuel using only technologies that do not produce any greenhouse gas emissions, and not lead to land use changes that increase greenhouse gases, then biofuels would not emit any net greenhouse gases over the lifecycle.

In practice, the lifecycle emissions of biofuels are affected by a number of factors. First, the feedstock used affects carbon emissions during the growing stage—for example, through fertilization. The most common feedstock used in the United States is corn. Brazilian ethanol is made from sugar cane. So-called "second generation" or "cellulosic" ethanol uses feedstocks that require little in the way of irrigation and fertilizer during the growing process, such as miscanthus and switchgrass. Second, the fuel used for generation of heat and electricity during the refining process affects emissions. Third, the calculation is affected by whether the coproducts from distilling, notably "distillery grains with solubles," are dried before being sold and whether the emissions from drying should be included, or treated as another product.

Fourth, the lifecycle emissions of corn-based biofuels are affected by the milling process. Corn ethanol is typically refined using either a dry or wet milling process.

Under wet milling, the corn is soaked in hot water and sulfurous acid. The starches from this mixture are then separated and fermented, leading to ethanol. Dry milling requires less energy and generates fewer greenhouse gas emissions, but does not yield as many coproducts as wet milling. Under dry milling, the corn is ground into flour and "cooked" along with enzymes, where yeast is added for fermentation. The ethanol is then separated from the liquid. The remaining component undergoes another process turning it into livestock feed.

Finally, and most difficult to estimate, increases in biofuel production can alter land use patterns elsewhere. For example, an increase in Brazilian sugar cane ethanol may reduce pasturelands and thus cause cattle farmers to cut down rainforest, which reduces the quantity of greenhouse gases sequestered by the rainforest. The influential paper by Searchinger et al. (2008) was the first to measure this factor, finding that once indirect land use effects are considered, corn-based ethanol can have nearly twice the greenhouse gas emissions of gasoline. A number of follow-up papers have found that while these effects may not be this large, they remain important. For example, Tyner, Taheripour, Zhuang, Birur, and Baldos (2010) argue that once changes in both international trade and crop yields are accounted for, corn ethanol results in fewer greenhouse gas emissions than gasoline, despite indirect land use changes.

How does the sum of these factors compare to the emissions of gasoline? The emissions of a gallon of gasoline over the entire lifecycle of its production depend on, amongst other things, the efficiency of the refinery and weight of the oil. A number of estimates exist. The California Air Resource Board (2011) estimated that an average gallon of California-refined gasoline generates 27.9 pounds of CO_2-equivalent greenhouse gas emissions. Roughly 19 pounds of this comes from the combustion of the gasoline, while the remainder comes from the emissions associated with refining, transporting, and so on. The 19 pounds figure may sound too high, given that a gallon of gasoline weighs roughly 6 pounds. The reason is that during the combustion process the carbon atoms in the gasoline, which have a molecular weight of 12, combine with 2 oxygen atoms from the atmosphere, each having a molecular weight of 16.

The California Air Resource Board (2011) also estimates that lifecycle emissions for a number of ethanol pathways lead to *highe*r greenhouse gas emissions than gasoline. For example, Midwest ethanol (shipped to California) produced using a wet mill process and coal for heating and electricity has 26 percent more greenhouse gas emissions than the average gasoline refined in California. In contrast, dry mill, wet "distillery grains with solubles" Californian ethanol which uses 80 percent natural gas and 20 percent biomass is predicted to have greenhouse emissions that are 19 percent below that of gasoline. Brazilian ethanol made from sugarcane has the lowest lifecycle emissions among those pathways analyzed in the California report. An Environmental Protection Agency (2009) report reaches similar conclusions. Dry mill ethanol made using coal has either 13 or 34 percent *more* emissions than gasoline. However, dry mill ethanol using biomass, a form of cellulosic ethanol, in a combined heat and power system has 26 or 47 percent fewer emissions.

In short, lifecycle analyses suggest that corn-based ethanol can play only a marginal role in reducing greenhouse gas emissions from the transportation sector. In contrast, cellulosic-based biofuels can potentially play a much larger role, although there remain technological obstacles to widespread mass production of ethanol at low cost from this source.

There are other natural limits to the impact of corn-based ethanol production in the United States as well. How much farmland would be required if America's cars were to run solely on E85, which is 85 percent ethanol and 15 percent gasoline? Well, gasoline usage in the United States is roughly 140 billion gallons per year, and it takes 128,500 acres of corn to produce 50 million gallons of ethanol (according to the FAQ at ⟨http://ethanol.org⟩). Given that ethanol has an energy content that is roughly 67 percent of gasoline, 140 billion gallons of our current fuel, which is roughly 5 percent ethanol, would equal roughly 190 billion gallons of E85. Thus, if the ethanol used corn as the feedstock, this would imply roughly 415 million acres of corn crop—but there is currently only 406 million acres of farmed land in the United States. In short, significant expansion of corn-based ethanol production is likely to require additional land, which unleashes environmental consequences discussed earlier. In addition, corn-based biofuels also compete with current uses of corn, which has implications for the worldwide price of corn and other substitute grains. Cellulosic biofuels, in contrast, offer a feedstock that will not compete with food products nearly as much, since these plants can be grown on marginal lands and without irrigation.

Large-scale substitution of ethanol for gasoline is limited in the short run because of the "blend wall"—the percentage of fuel that can be ethanol and safely burned in a vehicle designed to burn only gasoline. The Environmental Protection Agency recently ruled that vehicles of model year 2005, or newer, can safely burn fuel that is 15 percent ethanol. Vehicles older than this can burn E10. Flex-fuel vehicles, in contrast, can burn fuel that is up to 85 percent ethanol.

U.S. policymakers have adopted a variety of biofuel policies: performance standards, subsidies, and mandates. The Volumetric Ethanol Excise Tax Credit expired on December 31, 2011. The credit offered fuel blenders $0.45 tax credit per gallon of ethanol sold. Before this tax credit, ethanol received an implicit subsidy (relative to gasoline) as it was exempted from the federal fuel excise tax in 1978. The 2008 Farm Bill differentiated between corn-based and cellulosic ethanol, with cellulosic ethanol receiving a $0.91 per gallon tax credit, minus an applicable tax credit collected by the blender of the cellulosic ethanol.[3] Small ethanol producers—those with a capacity of less than 60 million gallons—received an additional 10 cents per gallon credit.

[3] These figures understate the subsidy level because they are on a per-gallon basis, not on a per-energy basis. As noted in the text, one gallon of ethanol has roughly 67 percent of the energy content of a gallon of gasoline, implying that it requires 1.48 gallons of ethanol to displace one gallon of gasoline. Therefore, on a per "gallon of gasoline equivalent" basis, corn-based ethanol received a 67 cents per gallon of gasoline equivalent subsidy; 81 cents for a small producer. Cellulosic ethanol received a $1.35 per gallon of gasoline equivalent subsidy, $1.49 per gallon of gasoline equivalent for small producers.

Similar subsidies existed for biodiesel. The Jobs Creation Act of 2004 established a $1-per-gallon tax credit for biodiesel created from "virgin" oil, defined as oil coming from animal fats or oilseed rather than recycled from cooking oil. Biodiesel from recycled oil receives a $0.50-per-gallon tax credit. These subsidies were extended under the Energy Policy Act of 2007, but also expired at the end of 2011.

The other major federal ethanol policy is mandates to use such fuels. The first Renewable Fuel Standard was adopted in 2005. The Energy Independence and Security Act of 2007 expanded this standard by calling for 36 billion gallons of biofuels—including 21 billion gallons of "advanced" biofuels by 2022, which are to have a lower greenhouse gas content than corn-based ethanol. Given how the Renewable Fuel Standard is implemented, ethanol prices reflect an implicit subsidy, while gasoline is priced as if it were taxed (Holland, Hughes, Knittel, and Parker 2011). A variety of state-level blend minimums and performance standards also exist.

Methanol is another alcohol that can be used as a liquid fuel. Methanol production is an established industry: methanol is used as a racing fuel, as an industrial chemical, and as a liquid fuel in some countries—especially China. Methanol can be produced from natural gas, coal, or biomass. In 2010, the United States consumed 1.8 billion gallons of methanol with world production totaling over 15 billion gallons (see statistics at ⟨http://methanol.org⟩), roughly on par with global ethanol production of 23 billion gallons in 2010 (see statistics at ⟨http://ethanolproducer.com⟩). In contrast to ethanol, most methanol consumption is not as a fuel, but as a chemical feedstock.

Methanol has three main advantages over corn-based ethanol. First, on a greenhouse gas basis, Delucchi (2005) estimates that methanol produced from natural gas has 11 percent lower greenhouse gas emissions than corn-based ethanol. However, he finds that methanol still has higher emissions than gasoline. Others find that the greenhouse gas emissions from methanol are roughly equivalent to gasoline (MIT 2011). Second, methanol is cheaper than gasoline, at least at current oil and gas prices. Methanex, the world's largest methanol producer, quotes current retail methanol prices in North America of $1.38 per gallon. Methanol has an even lower energy content than ethanol at roughly 53 percent of gasoline, so this implies a cost per gallon of gasoline equivalent of $2.51 per gallon (approximately, because of changes in engine efficiency), still cheaper than gasoline. Third, methanol doesn't rely on crops, eliminating the negative consequences associated with crop production.

Methanol also faces four disadvantages. First, methanol produced from natural gas cannot achieve the same reductions in greenhouse gas emissions as second-generation or cellulosic ethanol. Second, alcohols are generally more corrosive than gasoline, and methanol is even more corrosive than ethanol. For vehicles to run on ethanol or methanol, manufacturers must protect certain engine parts and rubber material from the fuels. Flex-fuel vehicles that can run on fuel that is as much as 85 percent methanol (M85) require a slightly larger investment, on the order of $200 per vehicle (MIT 2011). Third, as discussed above, methanol has an even lower energy content than ethanol, so a tank of gas wouldn't take you as far.[4] Finally, there are open questions as to how safe the drilling process is, or can be, for shale gas, including potential problems of methane leakage.

Given the recent discoveries of large shale gas deposits within North America, a compelling argument can be made that methanol, as a substitute for gasoline, should have the same support as corn-based ethanol. Methanol carries similar greenhouse gas reductions, if not larger, and is not petroleum based. The open issue is whether drilling for shale gas has fewer environmental repercussions than the land use implications of ethanol.

An alternative use for natural gas is in compressed natural gas (CNG) vehicles, which use internal combustion engines to burn natural gas stored at high pressures. Rood Werpy, Santine, Burnham, and Mintz (2010) summarize tailpipe emission comparisons of vehicles and find that compressed natural gas has emission reductions that are often above 20 percent, compared to gasoline, but often less than 10 percent when compared to diesel fuel. Moreover, long-run average costs on a gallon-of-gasoline equivalent are currently below gasoline: the U.S. Department of Energy reports national average prices of $2.09 for October 2011.

The drawbacks to CNG vehicles are similar to electric vehicles (discussed below). New infrastructure is needed for refueling with compressed natural gas. Refueling can take longer, especially if done at home: slow-fill home units can take over four hours. CNG vehicles have limited range, often the equivalent of about eight gallons of gasoline. CNG vehicles also have a higher upfront cost: the Honda Civic GX, a CNG vehicle, sells for, roughly, a $4,000 premium but has 27 percent less horsepower than a comparable gasoline-powered car. A thorough comparison of CNG and electric vehicles is beyond the scope of this paper, but again, given large natural gas reserves recently discovered, this would appear to be a worthwhile avenue for research. My read of the literature suggests that these drawbacks are not as severe with CNG vehicles as with electric vehicles, although the reduction in carbon emissions from CNG vehicles may also be less than if the electricity for a vehicle is generated in a low-carbon manner. Once the benefits of both greenhouse gas emissions and petroleum reductions are compared with the added costs, CNG vehicles might make more sense than electric vehicles.

Replacing the Internal Combustion Engine

Shifting away from the internal combustion engine to powering vehicles with electricity or with hydrogen is another way of reducing petroleum usage. Either approach could represent a reduction in the pollutants per unit of energy of the fuel and an increase in fuel economy—as measured by the energy required to travel

[4] Because of their lower vapor pressure, starting an engine in cold weather is more difficult when using ethanol and methanol (with ethanol having a lower vapor pressure compared to methanol), which may prompt consumers to use a lower blend of these fuels during the winter.

one mile. In terms of greenhouse gas emissions, all-electric or hydrogen vehicles have been viewed by some as the end game, since it is possible to generate either electricity or hydrogen in a carbon-free way—say, through solar or wind power. Ultimately, both technologies would probably use electric motors. It is possible to burn hydrogen directly in an internal combustion engine. BMW, for example, has a flex-fuel 7-series that can use both diesel and hydrogen. However, this forgoes the efficiency gain from electric motors, so most industry followers believe that if hydrogen were to penetrate the market it would do so through a fuel cell that powered an electric motor vehicle.[5]

The hurdle for both electricity and hydrogen technologies is, of course, cost. These costs can usefully be divided up into costs for vehicles, cost of the electricity or hydrogen itself, and infrastructure costs associated with reenergizing vehicles.

For a pure electric vehicle, battery technology still imposes some daunting constraints. While I am not aware of any studies detailing the required battery size as a function of key variables such as the vehicle weight and desired range, some rough calculations are possible. My personal communications with Yet-Ming Chiang of MIT suggest that a current mid-sized sedan, weighing about 3,000 pounds, requires roughly 300 watt-hours of battery capacity for every mile of range. This figure for a mid-sized sedan is roughly comparable to the 2011 Nissan Leaf, which weighs 3,354 pounds. The Leaf has a 24-kilowatt-hour battery pack and has a range rating of 73 miles from the Environmental Protection Agency, which translating to 328 watt hours per mile. A mid-sized sport-utility vehicle, weighing roughly 4,000 pounds, requires 425 watt hours for every mile of range. For a 200-mile range, which is significantly lower than current internal-combustion-based vehicles, the mid-sized sedan would require a 60-kilowatt-hour battery pack, while the mid-sized sport-utility vehicle would require a 85-kilowatt-hour battery pack. As a third point of reference, a 2011 Ford F-150 SuperCab weighs 5,500 pounds. If the relationship is roughly linear, a pickup truck of this size would require a 123-kilowatt-hour battery pack. I should note that I am ignoring the effects of the battery's weight, which have real consequences (Kromer and Heywood 2007). For example, the battery and control module for the Nissan Leaf weighs over 600 pounds.

A report from the National Research Council (2010) estimated current battery costs and projected future costs for plug-in hybrid vehicles. The committee set the most probable current cost for a battery at $875 per kilowatt hour, with $625 per kilowatt hour being an optimistic estimate. They project battery costs falling by 35 percent by 2020 and 45 percent by 2030. At these prices and assuming they scale up to the larger battery sizes required for all-electric vehicles, currently the battery alone for a mid-sized sedan with a range of 200 miles would cost between $38,000

[5] The efficiency of current electric motors is roughly 80 percent—meaning 80 percent of the energy in electricity goes to moving the vehicle, while current internal combustion engines are in the low 20 percent range. The theoretical bound on efficiency is roughly 30 percent for the internal combustion engine. For a reasonably accessible explanation, see Johnson (2003) at ⟨http://mb-soft.com/public2/engine.html⟩.

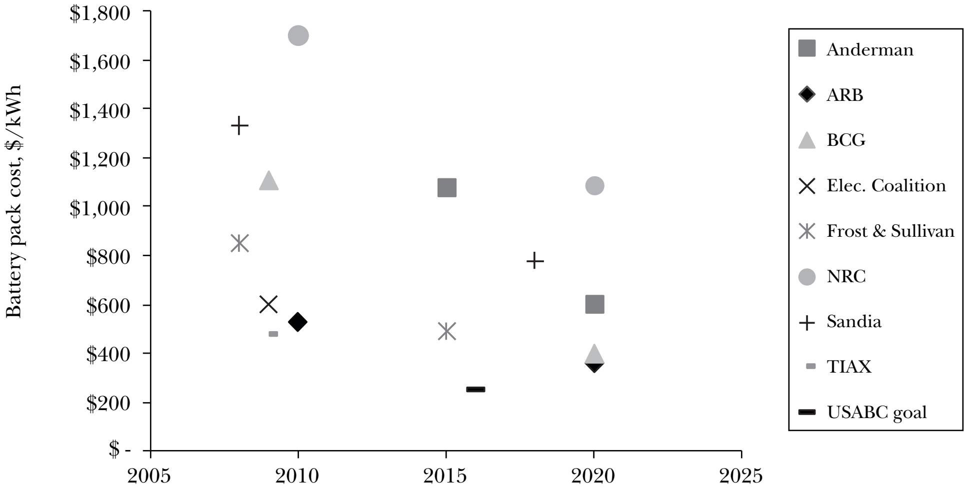
and \$50,000; the cost of a battery for the mid-sized sport-utility vehicle would be \$53,000–\$70,000; and a battery for the F-150 would cost between \$76,000 and \$101,000. The optimistic values in 2030 for battery costs alone would be \$21,000 for the sedan, \$29,000 for the sport-utility vehicle, and \$42,000 for the full-sized pick-up truck. The lower cost per mile of electric vehicles would offset these higher upfront costs to some extent. The sedan, for example, at average retail electricity rates would cost 3 cents per mile, compared to roughly 13 cents per mile at a gasoline price of \$4/gallon and a fuel economy of 30 MPG. However, these savings in operating costs are unlikely to outweigh the upfront costs at any reasonable discount rate (Anderson 2009).

While all-electric vehicles may not be cost competitive, vehicles that are partly propelled by electricity, such as hybrids or plug-in hybrids, may be. Hybrid and plug-in hybrid vehicles economize on battery costs because they use a higher share of the battery's capacity for typical driving patterns. Put another way, if a consumer could size the battery in an all-electric vehicle for each specific trip, all-electric vehicles might be cost competitive at current battery prices. To underline this point, Anderson (2009) calculates that a plug-in hybrid with a 10-mile range is cost competitive even at battery costs of nearly \$2,000 per kilowatt hour. Similar themes are echoed in the more comprehensive analysis of Michalek, Mikhail, Jaramillo, Samaras, Shiau, and Lave (2011).

The National Research Council (2010) battery cost estimates are somewhat controversial. The estimates accord well with the published cost estimates for the Nissan Leaf's battery of \$750 per kilowatt hour (Loveday 2010) and are within the range of estimates I have seen for the Chevrolet Volt's 16-kilowatt-hour battery pack (\$500–\$930) per kilowatt hour (Hall and Schoof 2011; Peterson 2011). However, a number of industry trade groups argue that their costs are too high (for example, Electrification Coalition 2009a; CalCars 2010). Better Place, a swappable electric vehicle battery company, has stated that they are purchasing batteries at \$400 per kilowatt hour. Other studies estimate much lower prices under hypothetical situations. For example, Nelson, Santini, and Barnes (2009) and Amjad, Neelakrishnan, and Rudramoorthy (2010) simulate battery costs as low as \$260 per kilowatt hour using engineering models of production. These results rely heavily on large scale economies and an assumption that plants operate 24 hours a day. Under these assumptions, costs fall by as much as an order of magnitude when production increases from 10,000 to 100,000 units per year. Figure 4 plots a number of battery cost estimates for different points in time, as well as the goal of the United States Advanced Battery Consortium, as summarized in the review article by Cheah and Heywood (2010); clearly, the estimates show a large dispersion in all years.

The true cost of batteries, both now and certainly in the future, is unresolved. But these calculations suggest that some major technological breakthrough may be needed for electric vehicles to play a large role in reducing oil consumption: either a much lower-cost battery, or technological breakthroughs that allow reductions in the size and/or weight of vehicles, perhaps through the use of polymer, aluminum, or composite body panels. However, technological breakthroughs reducing size

Figure 4
Battery Cost Estimates from the Literature
(as summarized in Cheah and Heywood 2010)



Source: Figure 4 reproduced from Cheah and Heywood (2010), "The Cost of Vehicle Electrification: A Literature Review."
Notes: Figure 4 plots a number of battery cost estimates for different points in time, as well as the goal of the United States Advanced Battery Consortium, as summarized in the review article Cheah and Heywood (2010). Cost estimates are from Anderman (2010), Air Resources Board (2009), Boston Consulting Group (2010) (BCG), Electrification Coalition (2009b), Frost & Sullivan (2009), National Research Council (2010), Ton et al. (2008) (Sandia), Barnett et al. (2009) (TIAX), and Pesaran, Markel, Tataria, and Howell (2007) (USABC). When a range is given in the original source, Cheah and Heywood plot the average. The USABC number is a goal, not a cost estimate.

and weight could also be applied to internal combustion engines and could thus have significant effects on oil use in that way—without leading to greater use of electric cars (Knittel 2011). Alternatively, the ranges of electric vehicles could end up being much shorter than we are accustomed to hearing about. Indeed, the battery-powered Nissan Leaf is rated at a range of 73 miles. Air conditioning or heating—because heat from the internal-combustion engine can no longer be used to heat the interior of the car—significantly reduces this range. *Car and Driver*'s road test for the Nissan Leaf finds an average range of 58 miles and discusses the effect of heating (Gluckman 2011).

Hydrogen vehicles also take advantage of the higher efficiency inherent in electric motors but generate their own electricity via a fuel cell. Support for hydrogen vehicles has significantly waned over the past decade, but pursuing the possibility of a hydrogen-fueled car remains a stated objective of the U.S. Department of Energy. Hydrogen vehicles use a fuel cell, which uses a "proton exchange membrane" to convert stored hydrogen, and oxygen from the surrounding air, into electricity; the by-product of this conversion is water. Fuel cells are cheaper than batteries and refueling could be much faster. (Although supporters of batteries sometimes argue that you could refuel quickly via a system of "swappable batteries.") At present, however, hydrogen refueling is not simple, with some stations requiring special suits and apparatus.

Detractors of hydrogen vehicles often point to the fact that they are far less efficient than electric vehicles on a "well to wheel" basis; that is, they take more total energy to travel one mile, because of the energy needed in making the hydrogen. However, the more relevant question is the relative cost of the two technologies. That is, if the added energy needed to produce hydrogen were free or low-cost to society, then the added inefficiency would not matter or would matter less. That is not to say hydrogen vehicles do, in fact, have lower costs. For hydrogen vehicles, the relevant costs are: the cost of the fuel cell, the cost of the high-pressure storage tank, the cost of hydrogen, and infrastructure costs.

The first main cost element for hydrogen-fueled cars are the fuel cells, which are currently quite expensive. A recent U.S. Department of Energy study (James, Kalinoski, and Baum 2011) estimates that the cost of fuel cells at the current fairly low production levels are roughly \$230 per kilowatt. To understand what this means for costs, the Chevy Volt has a 111-kilowatt electric motor, while the Nissan Leaf has a 80-kilowatt motor. The Volt's motor is equivalent to a 149-horsepower engine, which is about the amount of horsepower from a four-cylinder gasoline engine. Manufacturers appear to install fuel cells equivalent to the size of the motor, so the Volt would require a 111-kilowatt fuel cell at a cost over \$25,000. (There is a prototype Toyota Highlander FCV on loan to the University of California-Davis that combines a same-sized motor and fuel cell. The Honda FCX Clarity does so as well.) The alternative is to hybridize the vehicle by combining a fuel cell with a rechargeable battery back-up. Of course, electric motor and fuel cell combinations with horsepower levels comparable to larger vehicles would need to be correspondingly much larger.

As with some of the literature on battery costs, a number of papers on the future costs of fuel cells are built on assumptions of large scale economies. Using engineering-economic simulation models, the U.S. Department of Energy study assumes a scale economy elasticity of –0.2, and thus simulates that a fuel cell manufacturer producing 500,000 units per year could do so at an encouraging cost of \$51 per kilowatt (James, Kalinoski, and Baum 2011). Given the size of the possible gains from economies of scale and learning-by-doing, more studies along these lines would seem to be an important area for future research.

The second major cost component for a hydrogen vehicle is the storage tank. Hydrogen is ideally stored as a liquid under pressure because this has the highest energy density. BMW recently demonstrated a hydrogen vehicle with liquid storage. However, storing hydrogen as a liquid faces major obstacles, as the National Research Council (2004) study points out. For example, the liquid must be kept at –252 degrees Celsius, and the liquid storage tanks currently cost roughly \$500 per kilowatt hour of energy stored, with the "next generation" perhaps dropping the cost to roughly \$100 per kilowatt hour (Brunner 2006). Again using 60 kilowatt hours as a reasonable guideline for a mid-sized sedan that can travel 200 miles, the storage tank alone would cost $30,000 using current technology and $6,000 using the projected next-generation technology.

Thus, absent a major technological breakthrough in liquid storage, hydrogen is likely to be stored as a compressed gas, which either increases the space required for the storage tank or reduces the range of the vehicle (Ogden et al. 2011). Costs of compressed storage tanks, if produced at a large scale, might fall between $15 and $23 per kilowatt hour of energy (Ogden et al. 2011). Therefore, the storage tank for a 3,000-pound sedan with a range of 200 miles would cost between $900 and $1,400. However, gas storage tanks face durability issues, which are addressed by making the tanks larger. Indeed, the volume of a tank of this size is large enough that manufacturers are likely to design the vehicle around the tank (National Research Council 2004). However, if the estimated scale economies truly exist for both fuel cells and storage tanks, the combined cost of the fuel cell and storage tank for a hydrogen vehicle have the potential to be much cheaper than the battery required for an electric vehicle.

The third component is the cost of the hydrogen fuel itself, often quoted in terms of dollars per kilogram. A kilogram of hydrogen has roughly the same energy content as a gallon of gasoline, and given the increased efficiency of the electric drive-train, it can propel the vehicle roughly twice as far as a gallon of gasoline (for discussion, see National Research Council 2004, Appendix H). Here, too, the engineering literature suggests the possibility of large scale economies. Hydrogen can be produced in many ways, ranging from on-site production facilities to larger facilities where hydrogen is then shipped to refueling stations. Weinert and Lipman (2006) provide engineering cost estimates of the long-run average cost of hydrogen. Cost estimates vary considerably, but are as low as $4.90/kg. Accounting for the more efficient motors (and taxes on gasoline), this is roughly on par with current gasoline prices.

The current federal subsidy for electric vehicles is a tax credit of $2,500 plus $417 for each kilowatt hour of battery capacity in excess of 4 kilowatt hours, with a maximum tax credit of $7,500. Both the Nissan Leaf and Chevrolet Volt qualify for the maximum tax credit. The Toyota Prius Plug-in Hybrid, with a battery size of 4.4 kilowatt hours, qualifies for a $2,500 tax credit. There is also a federal tax credit for installation of charging equipment equal to 30 percent of the cost, with a maximum tax credit of $1,000 for residences and $50,000 for businesses (Belson 2011). A variety of state-level policies also exist with tax credits as high as $6,000 for qualifying vehicles (in Colorado).

One open question is whether, given the apparent need for technological breakthroughs for either electric or hydrogen vehicles, the funds used for these subsidies would be better served subsidizing research and development. The battery industry points to a number of potential “game changers,” such as lithium-air batteries and semi-solid flow cell batteries. Lithium-air batteries have a much higher energy density compared to the lithium-ion batteries presently used in the Leaf and Volt, leading to as much as five to ten times more energy for a given weight than lithium-ion batteries and twice the energy for a given size (Zyga 2011). However major hurdles exist. These batteries are prone to get "clogged" as lithium-oxide builds up in the battery, and therefore cannot be recharged as often as would be needed in a vehicle. Semi-solid cell batteries suspend the positive and negative electrodes in a liquid electrolyte (Chandler 2011). This not only has the potential for efficiency gains, but the battery can also, in principle, be "refueled" by draining the spent liquid and pumping in full-charged liquid. This battery structure is still in its infancy, however.

The Forgotten Channel: Reductions in Vehicle-Miles Traveled

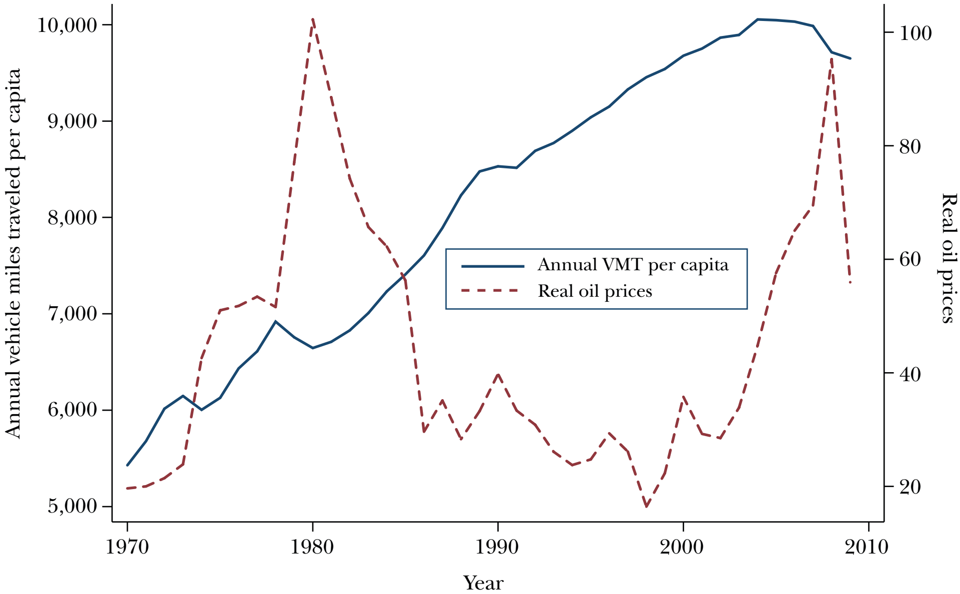
The final channel for reductions in oil consumption is reductions in vehicle-miles traveled. U.S. energy policy has largely ignored this channel. Indeed, policies like Corporate Average Fuel Economy standards and biofuel subsidies push in the opposite direction, in the sense that they reduce the marginal cost of driving an extra mile.

Figure 5 plots per capita vehicle-miles traveled in the U.S. from 1970 to 2009. The general trend upward is remarkable, with vehicle-miles traveled nearly doubling from 1970 to 2008. Remember that the figure shows *per capita* growth in vehicle-miles traveled, so that total growth in vehicle-miles traveled, including that attributable to population growth, would be even more striking. The figure also graphs real oil prices on the right-hand axis. The two price spikes in real oil prices—in the late 1970s and early 1980s, and in the last few years—are clearly correlated with a flattening out of vehicle-miles traveled. Conversely, the period of dropping oil prices over much of the intervening period is a time when vehicle-miles traveled soared. Of course, this connection is only illustrative: a full analysis of how the price of oil affects vehicle-miles traveled would need to make additional adjustments for changes in income, business cycles, and more. But more-detailed analyses do offer strong evidence that vehicle-miles traveled do respond to gasoline prices.

The response of vehicle-miles traveled to changes in gasoline prices varies, as one might expect, by the time frame for adjustment. The short run offers little scope for reductions in vehicle-miles traveled, and so the measured elasticity is likely to be small. For example, Small and van Dender (2007) estimate that one-month elasticity of vehicle-miles traveled to changes in price was –0.02 between 1997 and 2001, with a similarly calculated short-run elasticity of –0.05 from 1966 to 2001. In Hughes, Knittel, and Sperling (2008), my coauthors and I estimate the one-month elasticity for use of gasoline—which is largely driven by the one-month elasticity of vehicle-miles traveled. We find that the one-month elasticity in the 1970s was roughly –0.3, while it has fallen to roughly –0.07 in the 2000s.

In Knittel and Sandler (2011), we estimate an elasticity of vehicle-miles traveled over two years using observations on vehicle odometers in California's smog check program. We estimate an average elasticity of between –0.16 and –0.25 (see also Gillingham 2011, who finds similar estimates). More importantly, we find that the dirtiest quartile of vehicles in terms of their criteria pollutants are over four times

Figure 5
Vehicle-Miles Traveled per Capita from 1970 to 2009



Sources: Oil prices are taken from the Energy Information Administration. Vehicle-miles traveled (VMT) are from the Department of Transportation, ⟨http://www.fhwa.dot.gov/policyinformation/travel/tvt/history/⟩.

more responsive to changes in gasoline prices than the cleanest quartile, while dirtier vehicles in terms of their greenhouse gas emissions are roughly twice as sensitive. This increases the emission reductions resulting from a higher fuel price.

Long-run estimates of the elasticity of vehicle-miles traveled with respect to price are more difficult to identify, given that no sustained price increase exists in the data. The literature has thus focused on estimating partial adjustment models. This approach is necessarily imperfect, because if consumers believe the price change to be temporary, then the partial adjustment parameter leads to an underestimate of the long-run elasticity. Using a partial adjustment model, Small and van Dender (2008) estimate a long-run elasticity of vehicle-miles traveled with respect to price of –0.11 from 1997 to 2001 and –0.22 across their entire sample from 1966 to 2001. A number of earlier studies find roughly similar results to the short-run, medium-run, and long-run findings described here. Graham and Glaister (2002) and Dahl (1995) provide surveys.

Few existing policies seek reductions in vehicle-miles traveled, other than subsidies for public transit. Parry and Small (2009) present evidence that large public transit subsidies are welfare improving. The main benefit, however, arises through relieving congestion, not through a significant reduction in petroleum usage. Given the reluctance of policymakers to adopt Pigouvian taxes that would affect petroleum consumption, this approach to reducing petroleum use is likely to be underutilized.

Discussion: Pigouvian Taxes and Policy Choices

When economists are confronted with negative externalities, their trained reaction is that economic actors need an incentive to take the social costs of their actions into account in their decision at the margin. This outcome can happen through a Pigouvian tax, a system of tradeable permits, or a system of clarified property rights. Here, I use the Pigouvian tax—in this case, a tax on petroleum or greenhouse gas emissions that would include the value of the environmental and other externalities discussed at the start of this article—as a benchmark with which to compare other policy options for reducing U.S. petroleum use and greenhouse gas emissions.

Absent other market failures, it is clear that performance standards, such as Corporate Average Fuel Economy standards and Renewable Fuel Standards, will be less efficient than Pigouvian taxes in curbing gasoline consumption. Most basically, performance standards act as an implicit tax and subsidy program. Any product "better" than the standard is implicitly subsidized, while any product "worse" than the standard is implicitly taxed. In the case of fuel-related policies, such as the Renewable Fuel Standard, this implies an implicit subsidy for fuels that are "greener" but nonetheless emit greenhouse gases, driving a wedge between the Renewable Fuel Standard and the efficient policy.

Additional inefficiencies exist with respect to the Corporate Average Fuel Economy standard. At a basic level, it focuses on the wrong thing—fuel economy instead of total fuel consumption. CAFE only targets new vehicles and leads to subsidies for some vehicles. Finally, CAFE pushes consumers into more-fuel-efficient vehicles without changing the price of fuel, leading to more miles traveled. The empirical size of this last effect, known as "rebound," is a matter of ongoing research, but to the extent that rebound occurs, it necessarily leads to greater congestion, accidents, and criteria pollutant emissions relative to the status quo. These added externalities loom even larger as the first-best outcome would lead to reductions in vehicle-miles traveled, not increases.

The Corporate Average Fuel Efficiency standards have often been analyzed in comparison with a Pigouvian tax. For example, Kleit (2002) investigates the long-run effects of a 3 miles-per-gallon increase in the CAFE standard, and the gasoline tax that would achieve the same reduction in consumption of gasoline. He finds that the CAFE standard leads to a $3 billion per year social cost. An 11-cent gasoline tax achieves the same 5.1 billion gallon reduction in annual gasoline use at a social cost of $275 million. Austin and Dinan (2005) find similar results. They simulate the costs of a 3.8 miles-per-gallon increase in CAFE and the required gas tax that achieves the same gasoline reductions over the 14 years in which the change in CAFE becomes fully implemented. They find that CAFE is between 2.4 and 3.4 times more expensive than the equivalent gasoline tax. The average cost of an increase in CAFE of 3.8 miles per gallon, on a per ton of carbon dioxide abated basis, is between $33 and $40. Using the social cost of carbon estimates of Greenstone, Kopits, and Wolverton (2011), discussed earlier, this suggests that increases in CAFE of this magnitude either reduce or slightly increase welfare.

Jacobsen (2010) estimates the relative efficiency of Corporate Average Fuel Economy standards and gas taxes and also focuses on the differential impacts of CAFE across U.S., Asian, and European automakers. He finds that CAFE is over seven times more expensive than a gasoline tax that achieves the same reductions in greenhouse gas emissions over the first 10 years. When Jacobsen allows for manufacturers to change the technology included in the vehicles they offer, as opposed to manufactures having to meet the new standard with a different mix of the same vehicles they sold under the old standard, he finds that the cost of CAFE falls by over 60 percent, but is still over twice that of the cost of the gasoline tax without technology options. Even after allowing for technology adoption, Jacobsen estimates that the cost of a one-mile-per-gallon increase in CAFE is over $220 per ton of carbon dioxide saved, well above current estimates of the social cost of carbon.

A parallel line of research has compared subsidies and mandates for biofuels with a Pigouvian tax approach, again finding that the Pigouvian tax is much more cost effective. For example, Holland, Hughes, Knittel, and Parker (2011) simulate the relative efficiency of ethanol subsidies and the Renewable Fuel Standard compared to a Pigouvian carbon tax using feedstock-specific ethanol supply curves meant to represent cost conditions in 2020. They first simulate the greenhouse gas reductions from the current subsidies and Renewable Fuel Standard in 2020 and find reductions of 6.9 and 10.2 percent from subsidies and the Renewable Fuel Standard, respectively. They then calculate the required carbon tax that achieves a 10.2 percent reduction in emissions. Their results suggest that the social cost under subsidies is four times greater than the average social cost under the carbon tax, $82 per ton of carbon dioxide under the subsidies compared to $19 per ton for the carbon tax, despite the larger emission reductions under the tax. Similarly, the social cost under the Renewable Fuel Standard mandate is three times larger than the carbon tax, at $49 per ton of carbon dioxide.[6] There are also unintended consequences associated with the fuel-based policies. Holland et al. shows that land use patterns vary considerably across subsidy and mandate programs relative to the Pigouvian tax. If these land use changes exacerbate other negative externalities, such as fertilizer run-off or habitat loss, the inefficiency of these fuel-based programs will be understated.

[6] The authors also find, however, that the alternatives to the carbon tax may potentially yield large wins for low-populated counties—as high as $6,800 per capita per year. They then find that congressional voting on the Waxman–Markey cap-and-trade bill correlates with the simulated district gains and losses in predicted ways: congressional members whose districts gain the most under the Renewable Fuel Standard were less likely to vote for Waxman–Markey cap and trade, conditioning on the congressional member's political ideology, the district's per capita greenhouse gas emissions, whether the state is a coal mining state, and a variety of other potential determinants of voting behavior.

The existing work on both Corporate Average Fuel Economy standards and the Renewable Fuel Standard suggest that increasing these policies may actually *reduce* aggregate welfare, given current estimates of the social cost of greenhouse gases. However, the research described thus far leaves out a potential market failure that can make fuel efficiency standards and biofuels subsidies appear at least somewhat more attractive: that consumers are myopic in their preferences about fuel economy. Myopic consumers may be unwilling to invest a dollar at the time of purchase for a savings in present-discounted dollars sometime in the future. Consumer myopia implies that correcting prices with a Pigouvian tax is not sufficient to achieve the first-best outcome. This opens the door for additional policies *to complement* Pigouvian taxes. If consumers apply a discount rate that is larger than their true discount rate when purchasing vehicles, then policies that alter their choices or the relative prices of vehicles can, in principle, raise welfare.

Several recent papers have explored this topic, and the paper by Allcott and Greenstone in this symposium takes up what they call the "energy efficiency paradox" in more detail. The evidence appears mixed. As one example, Allcott and Wozny (2011) find that when purchasing a used vehicle, the average consumer discounts the future at a rate of 16 percent. In contrast, in Busse, Knittel, and Zettelmeyer (2011), we find no evidence of the energy paradox in the new car market, and implied discount rates are most often below 12 percent in the used car market. There is also a possibility that consumers may be acting optimally in the sense that a 16 percent discount rate is on par with their cost of capital, but the optimal discount rate from a policymaker's perspective may be much lower if, for example, society puts more weight on the welfare of future generations than a given consumer may choose to do.

If consumers are indeed myopic in their willingness to consider fuel efficiency, then second-best policies like CAFE standards counteract consumer myopia as they push demand to more fuel-efficient vehicles. They do not, however, eliminate the need for Pigouvian taxes. Given the potential importance of consumer discounting for optimal policy in both transportation and in other sectors, this open question warrants further research.

How large must myopia be for Corporate Average Fuel Economy standards to be welfare-improving in the absence of Pigouvian taxes? Fischer, Harrington, and Parry (2007) analyze the social cost of increases in CAFE standards under two assumptions regarding consumer myopia: a) consumers mistakenly inflate their discount rate by 14 percent over the true discount rate of 4.5 percent; and b) consumers care only about the fuel costs in the first three years of the vehicle's life, but their true discount rate is 4.5 percent. They consider the welfare implications of increasing CAFE in the presence of a variety of externalities, including local and global pollution, congestion costs, accident costs, and external costs associated with oil dependence. Their results suggest that CAFE standards are welfare-improving *only* in case "b." Increases in CAFE reduce aggregate welfare even when consumers undervalue the future by 14 percentage points. Welfare falls by roughly 20 cents per gallon of gasoline saved under this scenario. They do not consider the welfare costs of Pigouvian taxes under these different scenarios.

The results with respect to both Corporate Average Fuel Economy standards and the Renewable Fuel Standard underscore an important point: second-best (or third-best) policies need not be welfare-improving in the presence of negative externalities. Continued work that helps us better understand in what circumstances they do improve welfare, and the magnitude of other market failures, is needed.

Conclusion

A policy that puts a price on the externalities, like a carbon tax or cap-and-trade policy, would be desirable in addressing the externalities created by petroleum fuels in the U.S. economy. But both because such policies seem impractical for political reasons and because of the possibility of consumer myopia, there is potentially a role to be played by supplementary policies. Given the current state of technology, biofuels, electric vehicles, and hydrogen fuel-cell vehicles remain some years in the future. Their eventual commercial viability probably depends on a combination of technical breakthroughs, the emergence of economies of scale in production, continued high prices for gasoline, and policy.

Corporate Average Fuel Economy standards *can* play a useful role as a second-best policy, in pushing automobile technology developments that focus on fuel efficiency over horsepower and weight-adding ingredients. But just as the U.S. political system doesn't much like fuel taxes, it's worth noting that the political process found a way for the CAFE standards not to bind very much over the last few decades—by giving sports-utility vehicles and light trucks a lower standard, or even no standard at all. Furthermore, the literature calls into question whether increases in CAFE standards are welfare improving.

It will be interesting to see how the political system reacts if the significantly higher fuel economy standards planned for the next few years begin to bite for leading U.S. car manufacturers. It is worth considering some alternative second-best policies: for example, an open fuel standard that would require vehicles to be able to run on gasoline, ethanol, or methanol; gas guzzler–gas sipper "feebate" programs that mimic CAFE standards; or a vehicle-miles traveled tax. But ultimately, the single biggest influence on whether Americans reduce their consumption of petroleum-based fuels will probably be whether the forces of supply and demand in global markets that have kept oil prices relatively high since about 2005 continue to do so.

■ *I benefited from discussions with Hunt Allcott, Severin Borenstein, Joseph Doyle, Stephen Holland, Jonathan Hughes, Kenneth Gillingham, Donald MacKenzie, Joan Ogden, and Victor Stango.*

References

Air Resource Board, State of California. 2009. "Preliminary Assessment of the Need for Revisions to the Zero Emission Vehicle Regulation," Attachment A: "Status of ZEV Technology Commercialization (Technical Support Document)." November 25.

Allcott, Hunt, and Nathan Wozny. 2011. "Gasoline Prices, Fuel Economy, and the Energy Paradox." July 20. https://files.nyu.edu/ha32/public/research/Allcott%20and%20Wozny%202011%20-%20Gasoline%20Prices,%20Fuel%20Economy,%20and%20the%20Energy%20Paradox.pdf.

Amjad, Shaik, S. Neelakrishnan, and R. Rudramoorthy. 2009. "Review of Design Considerations and Technological Challenges for Successful Development and Deployment of Plug-in Hybrid Electric Vehicles." *Renewable and Sustainable Energy Reviews* 14(3): 1104–10.

An, Feng, Robert Earley, Lucia Green-Weiskel. 2011. "Global Overview on Fuel Efficiency and Motor Vehicle Emission Standards: Policy Options and Perspectives for International Cooperation." United Nations Background Paper No.3, CSD19/2011/BP3.

Anderman, M. 2010. "Emission Vehicle Staff Technical Report of 11/25/2009," including Attachment A: "Status of EV Technology Commercialization, Advanced Automotive Batteries." January 6.

Anderson, David L. 2009. "An Evaluation of Current and Future Costs for Lithium-Ion Batteries for Use in Electrified Vehicle Powertrains." http://hdl.handle.net/10161/1007.

Austin, David, and Terry Dinan. 2005. "Clearing the Air: The Costs and Consequences of Higher CAFE Standards and Increased Gasoline Taxes." *Journal of Environmental Economics and Management* 50(3): 562–82.

Barnett, Brian, David Ofer, Chris McCoy, Yong Yang, Todd Rhodes, Bookeun Oh, Mildred Hastbacka, Jane Rempel, and Suresh Sririramulu. 2009. "PHEV Battery Cost Assessment." 2009 DOE Merit Review, May (2009). Presentation. http://www1.eere.energy.gov/vehiclesandfuels/pdfs/merit_review_2009/energy_storage/es_02_barnett.pdf.

Belson, Ken. 2011. "Budget Shortfall Could Imperil Subsidies for Electric Vehicles." *New York Times,* November 17. http://wheels.blogs.nytimes.com/2011/11/10/budget-shortfall-could-imperil-subsidies-for-electric-vehicles/.

Boston Consulting Group. 2010. "Batteries for Electric Cars: Challenges, Opportunities, and the Outlook to 2020." http://www.bcg.com/documents/file36615.pdf.

Brunner, Tobias. 2006. "BMW CleanEnergy—Fuel Systems." Presentation at the CARB ZEV Technology Symposium, Sacramento, California. http://www.arb.ca.gov/msprog/zevprog/2006symposium/presentations/brunner.pdf.

Busse, Meghan R., Christopher R. Knittel, and Florian Zettelmeyer. 2011. "Pain at the Pump: The Effect of Gasoline Prices on New and Used Automobile Markets." May. http://web.mit.edu/knittel/www/papers/gaspaper_latest.pdf.

CalCars. 2010. "Rebuttals to Flawed National NAS Report—and Challenge to Battery Industry." January 15. http://www.calcars.org/calcars-news/1087.html.

California Air Resource Board. 2011. "Appendix A: Proposed Regulation Order Amendments to the Low Carbon Fuel Standard." January 6. http://www.arb.ca.gov/regact/2011/lcfs11/lcfsappa.pdf.

Chandler, David L. 2011. "New Battery Design Could Give Electric Vehicles a Jolt." *MIT News,* June 6. http://web.mit.edu/newsoffice/2011/flow-batteries-0606.html.

Cheah, Lynette, and John Heywood. 2010. "The Cost of Vehicle Electrification: A Literature Review." MIT Energy Initiative Symposium, April 8, 2010, Cambridge, MA. http://web.mit.edu/sloan-auto-lab/research/beforeh2/files/PHEV%20costs.pdf.

Currie, Janet, and Reed Walker. 2011. "Traffic Congestion and Infant Health: Evidence from EZ Pass." *American Economic Journal: Applied Economics* 3(1): 65–90.

Dahl, Carol A. 1995. "Demand for Transportation Fuels: A Survey of Demand Elasticities and Their Components." *Journal of Energy Literature* 1(2): 3–27.

Delucchi, Mark A. 2005. "A Multi-Country Analysis of Lifecycle Emissions from Transportation Fuels and Motor Vehicles." Institute of Transportation Studies Working Paper UCD-ITS-RR-05-10.

Electrification Coalition. 2009a. "Response to National Research Council Study: Transitions to Alternative Transportation Technologies: Plug-In Hybrid Electric Vehicles." http://www.electrificationcoalition.org/media/releases/response-national-research-council-study.

Electrification Coalition. 2009b. *Electrification Roadmap: Revolutionizing Transportation and Achieving Energy Security.* November. Electrification Coalition.

Environmental Protection Agency. 2009. "EPA Lifecycle Analysis of Greenhouse Gas Emissions from Renewable Fuels." EPA-420-F-09-024, May.

http://www.epa.gov/otaq/renewablefuels/420f09024.htm.

Fischer, Carolyn, Winston Harrington, and Ian W. H. Parry. 2007. "Should Automobile Fuel Economy Standards Be Tightened?" *Energy Journal* 28(4): 1–29.

Frost & Sullivan. 2009. "World Hybrid Electric and Electric Vehicle Lithium-ion Battery Market."

Gillingham, Kenneth. 2011. "How Do Consumers Respond to Gasoline Price Shocks? Heterogeneity in Vehicle Choice and Driving Behavior." http://www.yale.edu/gillingham/Gillingham_ConsumerResponseGasPrices.pdf.

Gluckman, David. 2011. "2011 Nissan Leaf SL—Long-Term Road Test Wrap-Up: Nissan's First Try at an EV Might Make a Good Second Car." *Car and Driver*, August. http://www.caranddriver.com/reviews/car/11q2/2011_nissan_leaf_sl-long-term_road_test_wrap-up.

Graham, Daniel J., and Stephen Glaister. 2002. "The Demand for Automobile Fuel: A Survey of Elasticities." *Journal of Transport Economics and Policy* 36(1): 1–26.

Greenstone, Michael, Elizabeth Kopits, and Ann Wolverton. 2011. "Estimating the Social Cost of Carbon for Use in U.S. Federal Rulemakings: A Summary and Interpretation." NBER Working Paper 16913.

Hall, Kevin G., and Renee Schoof. 2011. "Electric-Vehicle Technology Full Speed Ahead." McClatchy-Tribune News Service, April 10. http://www.bendbulletin.com/article/20110410/NEWS0107/104100318/.

Hughes, Jonathan E., Christopher R. Knittel, and Daniel Sperling. 2008. "Evidence of a Shift in the Short-Run Price Elasticity of Gasoline Demand." *Energy Journal* 29(1): 113–34.

Holland, Stephen P., Jonathan E. Hughes, Christopher R. Knittel, and Nathan Parker. 2011. "Some Inconvenient Truths about Climate Change Policy: The Distributional Impacts of Transportation Policy." NBER Working Paper 17390.

International Center for Technology Assessment (ICTA). 2005. "Gasoline Cost Externalities: Security and Protection Services." An Update to CTA's *Real Price of Gasoline* Report. January 25. http://www.icta.org/doc/RPG%20security%20update.pdf.

Jacobsen, Mark. 2010. "Evaluating U.S. Fuel Economy Standards in a Model with Producer and Household Heterogeneity." September. http://econ.ucsd.edu/~m3jacobs/Jacobsen_CAFE.pdf.

James, Brian D., Jeff Kalinoski, and Kevin Baum. 2011. "Manufacturing Cost Analysis of Fuel Cell Systems." Presentation at the 2011 DOE H_2 Program, May 12. http://www.hydrogen.energy.gov/pdfs/review11/fc018_james_2011_o.pdf.

Johnson, C. 2003. "Physics in an Automotive Engine." February (updated 9/17/2011). http://mb-soft.com/public2/engine.html.

Kanellos, Michael. 2010. "EV Batteries Plummet in Price: Down to $400 a kwH." greentechmedia, August 31. http://www.greentechmedia.com/articles/read/ev-batteries-dropping-rapidly-in-price/.

Klier, Thomas, and Joshua Linn. 2010. "The Price of Gasoline and the Demand for Fuel Economy: Evidence from Monthly Sales Data." *American Economic Journal: Economic Policy* 2(3): 134–53.

Kleit, Andrew N. 2002. "Impacts of Long-Range Increases in the Corporate Average Fuel Economy (CAFE) Standard." Working Paper 02-10, AEI-Brookings Joint Center for Regulatory Studies, October.

Knittel, Christopher R. 2011. "Automobiles on Steroids: Product Attribute Trade-Offs and Technological Progress in the Automobile Sector." *American Economic Review* 101(7): 3368–99.

Knittel, Christopher R., Douglas L. Miller, and Nicholas J. Sanders. 2011. "Caution Drivers! Children Present: Traffic, Pollution, and Infant Health." NBER Working Paper 17222.

Knittel, Christopher R., and Ryan Sandler. 2011. "Cleaning the Bathwater with the Baby: The Health Co-Benefits of Carbon Pricing in Transportation." NBER Working Paper 17390.

Kromer, Matthew A., and John B. Heywood. 2007. *Electric Powertrains: Opportunities and Challenges in the U.S. Light-Duty Vehicle Fleet.* Sloan Automotive Laboratory, LFEE 2007-03 RP, Sloan Automotive Laboratory, MIT.

Li, Shanjun, Christopher Timmons, and Roger van Haefan. 2009. "How Do Gasoline Prices Affect Fleet Fuel Economy?" *American Economic Journal: Economic Policy* 1(2): 113–37.

Loveday, Eric. 2010. "WSJ: Nissan Leaf Profitable by Year Three; Battery Cost Closer to $18,000." autobloggreen, May 15. http://green.autoblog.com/2010/05/15/nissan-leaf-profitable-by-year-three-battery-cost-closer-to-18/.

Michalek, Jeremy J., Mikhail Chester, Pauline Jaramillo, Constantine Samaras, Ching-Shin Norman Shiau, and Lester B. Lave. 2011. "Valuation of Plug-in Vehicle Life-Cycle Air Emissions and Oil Displacement Benefits." *PNAS* 108(40): 16554–58.

MIT. 2011. *Future of Natural Gas.* An Interdisciplinary MIT Study. Massachussetts Institute of Technology. http://web.mit.edu/mitei/research/studies/documents/natural-gas-2011/NaturalGas_Report.pdf.

National Research Council. 2004. *The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs.* National Academies Press.

National Research Council. 2010. *Transitions to Alternative Transportation Technologies: Plug-in Hybrid Electric Vehicles.* National Academies Press.

Nelson, Paul A., Danilo J. Santini, and James Barnes. 2009. "Factors Determining the Manufacturing Costs of Lithium-Ion Batteries for PHEVs." International Electric Vehicles Symposium EVS-24, Stavanger, Norway. http://www.transportation.anl.gov/pdfs/B/624.PDF.

Ogden, Joan, Christopher Yang, Joshua Cunningham, Nils Johnson, Xuping Li, Michael Nicholas, Nathan Parker, and Yongling Sun. 2011. Chap. 3 in "The Hydrogen Fuel Pathway." In *Sustainable Transportation Energy Pathways,* edited by Joan Ogden and Lorraine Anderson. Institute of Transporation Studies, University of California, Davis.

Parry, Ian W., and Kenneth A. Small. "Should Urban Transit Subsidies Be Reduced?" *American Economic Review* 99(3): 700–724, 2009.

Pesaran, Ahmad, T. Markel, H. Tataria, D. Howell. 2007. "Battery Requirements for Plug-In Hybrid Electric Vehicles: Analysis and Rationale." Presented at the 23rd International Electric Vehicle Symposium, Anaheim, California, December 2–5. www.nrel.gov/docs/fy09osti/42240.pdf.

Peterson, Thane. "Review: 2011 Chevrolet Volt." *Bloomberg Businessweek,* May 3. http://www.businessweek.com/lifestyle/content/may2011/bw2011052_182710.htm.

Roland, Neil. 2011. "The New Math: How Obama's Compromises Make New CAFE a Smaller Leap."*Automotive News,* May 8.

Rood Werpy, M., D. Santine, A. Burnham, and M. Mintz. 2010. "Natural Gas Vehicles: Status, Barriers, and Opportunities." Argonne National Laboratory Working Paper ANL/ESD/10-4, August.

Schimek, Paul. 1996. "Gasoline and Travel Demand Models Using Time Series and Cross-Section Data from the United States." *Transportation Research Record* 1558: 83–89. doi: 10.3141/1558-12.

Schipper, Lee. 2006. "Automobile Fuel Economy and CO_2 Emissions in Industrialized Countries: Troubling Trends through 2005/2006." EMBARQ Working Paper. http://www.embarq.org/sites/default/files/automobile_fuel.pdf.

Searchinger, Timothy, Ralph Heimlich, R. A. Houghton, Fengxia Dong, Amani Elobeid, Jacinto Fabiosa, Simla Tokgoz, Dermot Hayes, and Tun-Hsiang Yu. 2008. "Use of U.S. Croplands for Biofuels Increases Greenhouse Gases through Emissions from Land-Use Change." *Science* 319(5867): 1238–40.

Small, Kenneth A., and Kurt van Dender. 2007. "Fuel Efficiency and Motor Vehicle Travel: The Declining Rebound Effect." *Energy Journal* 28(1): 25–51.

Tol, Richard S. J. 2008. "The Social Cost of Carbon: Trends, Outliers and Catastrophes." *Economics,* August 12, vol. 2. no. 2008-25.

Ton, Dan T., Charles Hanley, Georgianne H. Peek, and John D. Boyles. 2008. "Solar Energy Grid Integration Systems: Energy Storage." Sandia Report SAND2008-4247, July.

Tyner, Wallace E., Farzad Taheripour, Qianlai Zhuang, Dileep Birur, and Uris Baldos. 2010. "Land Use Changes and Consequent CO_2 Emissions due to U.S. Corn Ethanol Production: A Comprehensive Analysis." Final Report. Available at Argonne National Laboratory website: http://www.transportation.anl.gov/pdfs/MC/625.PDF.

U.S. Department of Energy. 2011. "Clean Cities Alternative Fuel Price Report." October. http://www.afdc.energy.gov/afdc/pdfs/Oct_2011_AFPR.pdf.

U.S. Energy Information Administration. 2007. "Fuel Economy Standards for New Light Trucks." http://205.254.135.24/oiaf/aeo/otheranalysis/fuel.html.

U.S. Environmental Protection Agency. 2010. "EPA and NHTSA Finalize Historic National Program to Reduce Greenhouse Gases and Improve Fuel Economy for Cars and Trucks." EPA-420-F-10-014, April. http://www.epa.gov/oms/climate/regulations/420f10014.htm.

Weinert, Jonathan X., and Timothy E. Lipman. 2006. "An Assessment of the Near-Term Costs of Hydrogen Refueling Stations, and Station Components." University of California, Davis, ITS Working Paper, UCD-ITS-RR-06-03, January.

Zyga, Lisa. 2011. "Lithium-Air Batteries' High Energy Density Could Extend Range of Electric Vehicles." Physorg.com. February 1. http://www.physorg.com/news/2011-02-lithium-air-batteries-high-energy-density.html.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 119–138

How Will Energy Demand Develop in the Developing World?†

Catherine Wolfram, Orie Shelef, and Paul Gertler

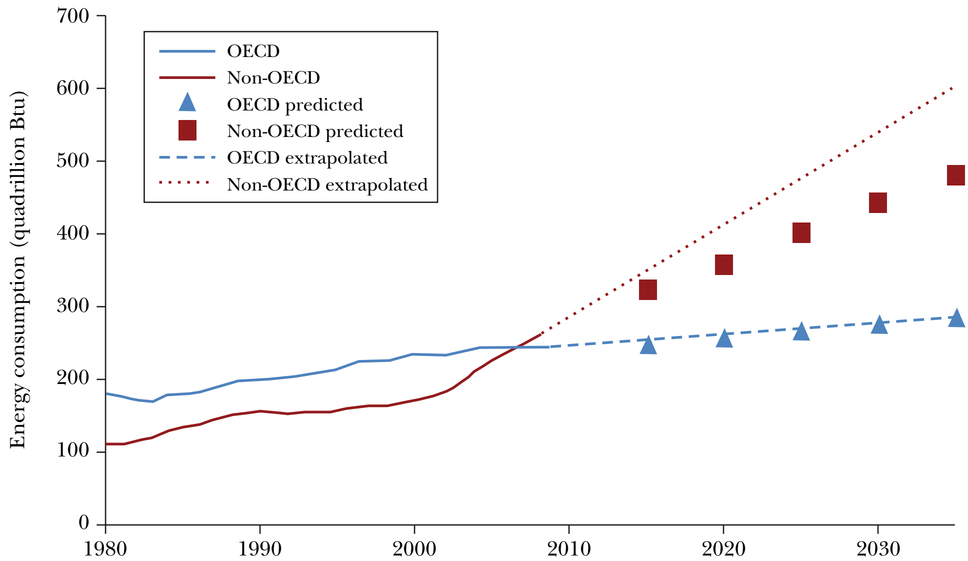
Over the next 25 to 30 years, nearly all of the growth in energy demand, fossil fuel use, associated local pollution, and greenhouse gas emissions is forecast to come from the developing world. The U.S. Energy Information Administration (2010a, table 1) reports that energy consumption in OECD and non-OECD countries was roughly equal in 2007, but from 2007 to 2035, it forecasts that energy consumption in OECD countries will grow by 14 percent, while energy consumption in non-OECD countries will grow by 84 percent.

This paper argues that the world's poor and near-poor will play a major role in driving medium-run growth in energy consumption. As the world economy expands and poor households' incomes rise, they are likely to get connected to the electricity grid, gain access to good roads, and purchase energy-using assets like appliances and vehicles for the first time. The energy needed to manufacture and use these new assets is likely to constitute a large portion of the growth in the demand for energy in the medium term. Also, refrigerators and cars are long-lived durable goods, so increases along the extensive margin driven by first-time purchases of these assets will have substantial consequences for energy consumption and greenhouse gas emissions for some time.

■ *Catherine Wolfram is Associate Professor and Co-Director of the Energy Institute, Orie Shelef is a doctoral candidate, and Paul Gertler is Li Ka Shing Professor, all at the Haas School of Business, University of California at Berkeley, Berkeley, California. Wolfram and Gertler are also Research Associates, National Bureau of Economic Research, Cambridge, Massachusetts. Their e-mail addresses are ⟨wolfram@haas.berkeley.edu⟩, ⟨orie_shelef@haas.berkeley.edu⟩, and ⟨gertler@haas.berkeley.edu⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.1.119.
doi=10.1257/jep.26.1.119

Figure 1
Energy Consumption in the Developing and Developed Worlds: Actual and Forecast



Source: U.S. Energy Information Administration, *International Energy Annual 2006* and *International Energy Outlook 2011.*

More broadly, the relationship between economic growth and energy consumption in the developing world has been, and is likely to continue to be, heavily influenced by the extent to which that growth is "pro-poor"—that is, by the extent to which growth improves the economic condition of those previously living in poverty. Further, we argue that the current forecasts for energy demand in the developing world may be understated, because they do not accurately capture increased demand along the extensive margin.

Trends in Energy Use in Developing Countries

Figure 1 plots total energy use in the developed and developing worlds, using actual numbers from 1980 to 2008 and Energy Information Administration projections out to 2035. Total energy includes consumption from the residential, industrial, commercial, and transportation sectors from different primary sources, including petroleum, natural gas, and coal, and electricity generated from the same fossil fuels, as well as from nuclear, hydro, geothermal, solar, wind, and biomass. The solid lines, plotting the actual numbers, demonstrate that energy consumption in the developing world has overtaken the developed world. In part, this reflects accelerating growth in developing world energy consumption in the most recent ten to fifteen years.

The squares and triangles plot the Energy Information Administration's forecasts for how energy use will evolve. Clearly, much of the growth is expected to be in the developing world. By 2035, developing world demand will almost double developed world demand. Other organizations offer similar forecasts. The International Energy Agency (2010, table 2.3) projects compound annual growth in non-OECD countries' energy demand of 2.0 percent through 2030, compared with the U.S. Energy Information Administration's (2010a, table 1) projected compound annual growth of 2.2 percent through 2035.

Macro-level energy-demand forecasts are difficult to develop and inherently uncertain, but they are also critical for future planning by energy producers, firms that rely on energy as an input, and scientists and others interested in understanding the possible range of increased greenhouse gases in the atmosphere. Underestimates can lead to underinvestment in energy production capacity, shortages and price spikes, and misunderstandings about both total future greenhouse gas emissions as well as country-specific emissions trajectories.

It is worth considering several benchmarks for the Energy Information Administration's projections. The dashed lines in Figure 1 provide one reference point. They represent extrapolations of the linear trends from 2002 to 2008 in the developing and developed worlds. These indicate that the Energy Information Administration is projecting that the developed world will follow the linear trend while the developing world will grow more slowly than it has recently.

As a second reference point, consider a crude forecasting exercise. Dividing 2008 energy consumption, as reflected in Figure 1, by total population, the developed world consumed 202 million BTU per person and the developing world consumed 47. If every person in the developing world increased energy use to the 2008 level of the developed world, developing world energy use would quadruple. Though this calculation is substantially higher than the Energy Information Administration's forecast, it could underestimate the potential growth in energy demand for several reasons. For one, it ignores population growth. It also likely understates the current gap between energy use in the developed and developing worlds, because a certain share of the energy currently used in the developing world is to produce goods for export to developed world consumers. On the other hand, this exercise could also overestimate energy use, because it is highly unlikely that every person in the developing world will achieve the level of consumption of the average person in the developed world by 2035. Also, it is possible that future energy use will become more efficient over time.

As this projection suggests, however, understanding the growth in energy demand involves understanding the process by which developing world consumers evolve into developed world consumers, to which we turn next.[1]

[1] Several papers have used country-level data and time series models to test for convergence in CO_2 emissions (for example, Strazicich and List 2003; Aldy 2006; Barassi, Cole, and Elliot 2011). Results are mixed, but generally suggest some convergence in emission rates among OECD countries but not worldwide.

Household-Level Energy Use

The economics profession has recognized for some time that household energy demand is driven as much by ownership of energy-using assets such as refrigerators and vehicles as by their usage (Dubin and McFadden 1984). Developing countries have recently experienced tremendous growth in sales of these energy-using assets. For example, in India there were 600,000 new vehicles sold in 2003, compared to 2,300,000 new vehicles sold in 2010 (Chugh 2011). Similarly, in urban China there were eight air conditioning units for every 100 households in 1995; by 2009, there were 106 units for every 100 households (Auffhammer 2011). Such patterns represent a general trend seen throughout the developing world: as households rise out of poverty and enter the middle class, they purchase new assets, many of which use substantial amounts of energy.

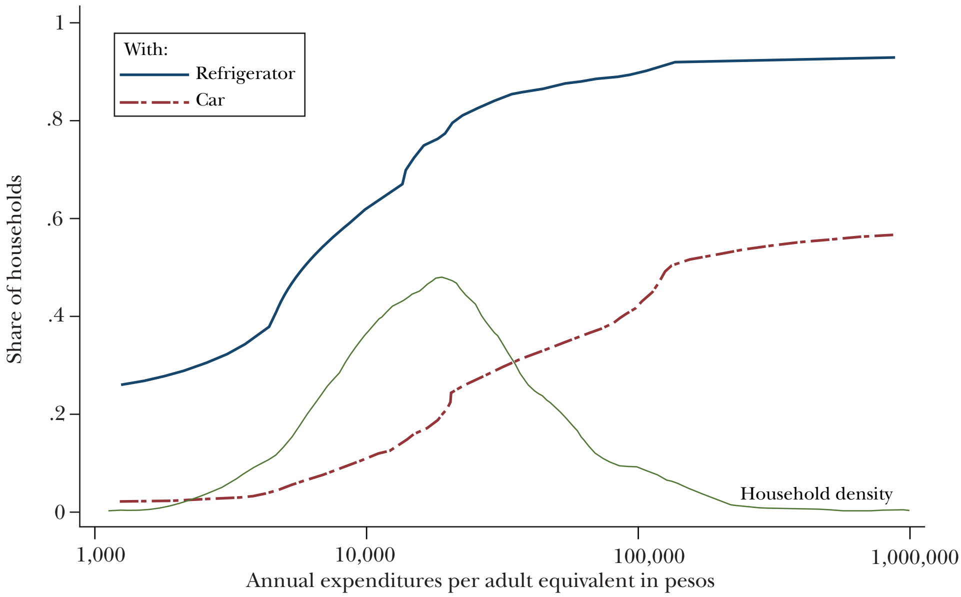
Household Appliance and Vehicle Holdings by Income Level: Cross-Sectional Evidence

To document this trend, we have assembled household-level survey data on appliance and vehicle ownership in several large developing countries, including China, India, Brazil, Indonesia, Mexico, and most of sub-Saharan Africa. The data are described more fully in the online Appendix available with this paper at ⟨http://e-jep.org⟩. As a generic category, appliances can include fans, air conditioners, washing machines, water heaters, blenders, irons, televisions, and more. Vehicles can include scooters, motorcycles, cars, and trucks. Our analysis focuses on refrigerators and cars, which are the two assets most consistently included in household-level surveys. Also, refrigerators are one of the first assets, after a television that a typical low-income household acquires. Moreover, the basic household decision making that drives refrigerator and car purchases applies to a range of other expensive, durable, energy-using assets. Finally, refrigerators and cars account for a significant share of developing world residential energy consumption. For example, refrigerators in China account for nearly 30 percent of residential electricity demand, or 15 percent of total residential energy demand.[2]

For most countries, we also use the surveys to measure each household's annual consumption expenditures as a measure of its overall well-being. We use household expenditures and not income for two reasons. First, data on expenditures are more reliable than data on income, particularly for households at the low end of the distribution who may have substantial informal and nonmonetary income sources. Second, if consumers smooth consumption either over their lifetime or across households (within extended families, for example), expenditures provide a better representation of household well-being. For Brazil, we use household

[2] Household refrigerators in China consumed approximately 145 terawatt-hours (TWh) of electricity in 2009 (Zhou et al. 2011; personal communication) while total residential electricity was 490 TWh (National Bureau of Statistics 2010). Electricity comprises 50 percent of residential energy consumption (Ni 2009).

Figure 2
Ownership Rates by Consumption Level: Mexican Refrigerators and Cars, 2000



Source: 2000 Mexico *Encuesta Nacional de Ingreso y Gasto de los Hogares* (a survey). See data Appendix for details.
Note: To measure the number of "adult equivalents" per household, we added household members 12 or older plus .5 times household members under 12.

income, because that country's survey does not include a comprehensive measure of consumption.

As an example, Figure 2 plots the relationship between annual household expenditures and refrigerator and car ownership in Mexico in 2000.[3] The distinctive S-shaped pattern that emerges is pervasive across our datasets and has been identified by others plotting country-level vehicle ownership against GDP (Dargay, Gately, and Sommer 2007; Dargay and Gately 1999). The figure also plots the density of households in Mexico by annual household expenditure level. For the approximately 10 percent of the Mexican households that consume less than 8,000 pesos (equivalent to roughly $800) per person per year, refrigerator and especially car ownership are uncommon. Also, the relatively flat slope of both the refrigerator and the car lines on the left quarter of the figure suggest that there are not large differences in refrigerator or car holdings even for low-income households whose total expenditures differ by a factor of two or more. Both curves reach an inflection

[3] Data from 1992 to 2008 show similar patterns, although there has been rapid growth in refrigerator ownership at the low end of the income distribution, particularly for refrigerators, so the S-shape is slightly more compressed in recent years.

point and become much more steeply sloped near the mode of the expenditure distribution, where most of the households are. At the high end, the curves flatten out, suggesting that above a certain threshold, cross-household differences in expenditures, measuring lifetime income, do not drive refrigerator or car purchase decisions. Refrigerators in Mexico in particular appear to reach a saturation point. The main focus of our analysis is the inflection point to the left of the graph. While the data in Figure 2 are cross-sectional and should not be interpreted as causal, they do suggest that as households rise out of poverty, many of them become first-time purchasers of energy-using assets.

Conceptual Framework

Early work such as Farrell (1954) assumed an S-shaped relationship between income and share of households with an asset, in a model based on a log-normal distribution of "acquisition thresholds." In Gertler, Shelef, Wolfram, and Fuchs (2011), we derive the S-shaped curve by modeling the appliance or vehicle acquisition decision and adding features that we argue are especially relevant in the developing world. The basic logic is straightforward. Households face a choice between consuming a divisible good with decreasing marginal utility (such as food) and an indivisible appliance that provides a fixed utility. As household income increases, the utility from increased consumption of the divisible good declines and, the probability that the household's utility from the appliance exceeds the utility from forgone food increases. Under reasonable assumptions on the distribution of appliance or vehicle valuations, this generates an S-shaped ownership curve.

Further, most energy-using assets are expensive and most low-income households in the developing world are credit-constrained.[4] A household does not a make a period-by-period choice of whether to own an asset effectively by renting it, as is assumed in much of the developed-market literature. Instead, the household must save to acquire the asset, which delays the asset acquisition to a higher income than would be suggested by the rental model. Because lower-income households are less able to self-finance, this delay is bigger at lower income levels and the resulting S-curve becomes steeper. Also, if households are self-financing through savings, we show that growth in income, and not just current income, will affect asset acquisition (Gertler, Shelef, Wolfram, and Fuchs 2011).

Household Appliance Holdings by Income Level: Evidence over Time

Though an S-shaped ownership curve is reflected in the cross-section and consistent with theory, it is possible that there are other relevant variables that are correlated with both income and a household's value for a refrigerator or car, in which case simple plots like Figure 2 are misleading. One obvious candidate is access to electricity or roads. Electrification rates in Mexico are about 98 percent, so

[4] A number of papers have noted the credit constraints on low-income households in the developing world, including Banerjee and Newman (1993), Aghion and Bolton (1997), Lindh and Ohlsson (1998), Lloyd-Ellis and Bernhardt (2000), Banerjee (2004), and de Mel, McKenzie, and Woodruff (2008).

that is unlikely to explain the refrigerator patterns. Still, the inflection point could reflect households from different regions, with the ones on the very left of the curve from cultures that place low value on refrigerated food or where communal assets are often shared. However, several additional facts suggest that such explanations play a relatively minor role compared to income.

One piece of evidence against the omitted variable hypothesis comes from our own work in Mexico. We use the income variation created by the conditional cash transfer program *Oportunidades* to examine appliance acquisition patterns (Gertler, Shelef, Wolfram, and Fuchs 2011). Each household's transfer payment was set according to a nonlinear function of the age and gender of children. Also, villages were randomly assigned to either the treatment or control group, and households in treated villages began receiving transfers 18 months before households in the control villages. We argue that the resulting cross- and within-household variation in transfers is exogenous to other factors likely to determine appliance acquisitions. In that paper, we develop three predictions on asset acquisition, which follow from the interplay between decreasing marginal utility from other goods and credit constraints, and find evidence consistent with each of them.

To confirm that the relationship we found among poor Mexicans matches a broader swath of the developing world, we use data from household expenditure surveys from several large developing countries. We use what Deaton (1986) calls a "pseudo-panel data approach," in which we analyze households at the same quartile of the income distribution over time. This approach controls for any time-invariant omitted variables that are correlated with both asset valuations and location in the income distribution. For simplicity, in Table 1 we group the middle two quartiles. For each country, the "Baseline ownership" column reports the share of each quartile owning a refrigerator or car in the earliest year for which we have data. The "Annual acquisitions" column reports the annual share of households that acquired refrigerators or cars between the baseline and final year, based on a linear interpolation. In other words, by 2009, 84 percent of the households in the bottom expenditure quartile in Brazil owned refrigerators, reflecting a 44 percentage point increase over the 17-year period (or a 2.6 percentage point annual increase). The data reported in the table reflect almost 4 billion people, a sizable share of the developing world.

The first three rows of Table 1 clearly suggest that in Brazil, Mexico, and urban China, growth in refrigerator ownership at the lower end of the income distribution was faster than growth at the higher end, which is consistent with the S-shaped pattern. The next row, for rural China, depicts slower growth at the low end of the income distribution. Poor households in rural China, however, were much poorer than their counterparts in Brazil, Mexico, or urban China and began in the 1990s with many fewer refrigerators per household. This suggests that most poor rural Chinese households were to the left of the inflection point in the S-curve. Households in the middle and upper parts of the distribution acquired refrigerators more rapidly, consistent with being located on the middle of the S-curve. Refrigerator acquisitions in India and Indonesia follow the same pattern as rural China, as do car acquisitions in urban China, India, and Indonesia. Car acquisitions in Mexico appear to be slower

Table 1
Growth in Refrigerator and Car Ownership by Income Quartile

	Bottom quartile		*Middle quartiles*		*Top quartile*	
	Baseline ownership	***Annual acquisitions***	*Baseline ownership*	***Annual acquisitions***	*Baseline ownership*	***Annual acquisitions***
Refrigerators						
Brazil (1992–2009)	40%	**2.6%**	76%	**1.1%**	96%	**0.2%**
Mexico (1996–2008)	32%	**2.9%**	70%	**1.4%**	91%	**0.4%**
China - urban (1995–2002)	41%	**2.9%**	55%	**3.0%**	87%	**0.6%**
China - rural (1995–2002)	1%	**1.1%**	4%	**1.8%**	15%	**2.5%**
India (2000–2007)	<0.5%	**<0.05%**	3%	**0.4%**	30%	**2.7%**
Indonesia (1999–2004)	2%	**<0.05%**	9%	**0.1%**	40%	**1.9%**
Sub-Saharan Africa (1994–2005)[a]	—	—	8%	**0.3%**	—	—
Cars						
Mexico (1996–2008)	2%	**0.5%**	16%	**0.9%**	49%	**0.3%**
China - urban (1995–2002)	<0.5%	**<0.05%**	<0.5%	**<0.05%**	<0.5%	**0.2%**
India (2000–2007)	<0.5%	**<0.05%**	<0.5%	**<0.05%**	4%	**0.4%**
Indonesia (1999–2004)	1%	**<0.05%**	3%	**<0.05%**	14%	**0.3%**
Sub-Saharan Africa (1994–2005)[a]	—	—	4%	**0.1%**	—	—

Sources: Data sources are described in the online data Appendix available with this paper at ⟨http://e-jep.org⟩.
Notes: For each country, the "Baseline ownership" columns reports the share of each quartile owning a refrigerator or car in the earliest year for which we have data. The "Annual aquisitions" column reports the annual share of households that acquired refrigerators or cars between the baseline and final year, based on a linear interpolation. Standard errors are suggested by the number of observations in each survey, which are reported in the online Appendix. Except in Brazil, quartiles are based on total surveyed consumption per adult equivalent (per capita in India). For Brazil, quartiles are based on stated income per adult equivalent instead of consumption.
[a]Sub-Saharan Africa is not divided by quartile, and all households are aggregated in the middle columns.

for the upper income quartile than for the middle quartile, although high-income Mexicans started from a higher base than any of the other countries.

The data we have for Africa come from Demographic and Health Surveys (DHS) performed by the USAID, which do not include data on household income or consumption, so we cannot break out the ownership rates by income quartile. We aggregate baseline ownership and growth rates for every country in sub-Saharan Africa with more than one DHS survey that identifies asset ownership. In all, the sub-Saharan Africa data cover approximately 500 million people. The numbers suggest that very few households in the region own refrigerators or cars, and growth in ownership has been languid. The big step toward ownership of these assets is yet to arrive in most of sub-Saharan Africa.

Household Energy Expenditures by Income Level

The patterns for appliances and vehicles strongly suggest that energy use will grow more quickly for households coming out of poverty than for households further up the income distribution. To verify this interpretation, we examined patterns in household energy expenditures by consumption quartile. Unfortunately, very few of the household expenditure surveys other than Mexico consistently ask households about electricity or other forms of energy expenditures. However, the data for Mexico clearly show that per capita electricity expenditures grew much more quickly for the average household in the lowest income quartile than for the average household in the middle or top quartile. Between 1996 and 2008, electricity expenditures doubled for households at the low end and only grew by 50 percent for households with higher expenditures.

These results are not driven by changes in the number of household members across quartile: similar patterns emerge if we look at expenditures per household. In addition, these trends do not appear to be driven by changes in relative prices. Over the early part of the sample, through 2002, prices did rise more slowly for high-volume users than for low-volume users, and use is correlated with income. In the later part of the sample, however, prices rose more slowly for low-volume users, and this is the period when expenditures deviated most dramatically between the two groups. This suggests that the differences across quartiles in the later part of the sample if anything understate different growth rates in consumption.

Electrification

Progress to Date

While the household-level data suggest the potential for rapid adoption of energy-using assets as incomes rise in the developing world, several important decisions are out of households' control. The utility a household derives from a car or refrigerator depends heavily—in the case of a refrigerator almost entirely—on whether the household has access to good roads or a reliable electricity source. This section will focus on electrification rates across countries, although it would be interesting to perform a similar analysis for roads or other components of the energy infrastructure, including natural gas pipelines and gasoline refueling stations.

More than one in five people worldwide—approximately 1.5 billion people—live without electricity in their homes. Understanding where these people are, and the process by which they gain access to modern energy, is crucial to understanding the growth in energy consumption in the developing world.

Table 2 documents the ten countries where the largest numbers of people live without electricity. Though methodologies for determining electrification vary by country, the data appear to capture all households with access to electricity whether or not they pay for it.[5] Sub-Saharan Africa is relatively underrepresented

[5] Note that simple measures of electrification rates abstract from variations in power quality, such as frequent blackouts. McRae (2009) explains how public policies that subsidize electric companies for

Table 2
In Which Countries Do the Most People Live without Electricity?
(top-ten countries by population)

	Electrification rate	*Number of people without electricity (millions)*
India	65%	404.5
Bangladesh	41%	94.9
Indonesia	65%	81.1
Nigeria	47%	80.6
Pakistan	58%	70.4
Ethiopia	15%	68.7
Democratic Republic of Congo	11%	57.0
Myanmar	13%	42.8
Tanzania	12%	36.8

Source: International Energy Agency, *World Energy Outlook 2009.*
Notes: The *World Energy Outlook 2011* lists higher electrification rates for several of these countries. The most dramatic change is for India, with a 75% electrification rate and 288 million people without electricity. The difference appears to be due to a new data collection methodology and highlights the uncertainty surrounding energy use in the developing world.

in this ranking because it is divided into many smaller countries. Across the region, the majority of the people (71 percent) live without electricity. This is certainly related to the low refrigerator penetration we discussed above. It is also notable that Brazil and China are *not* listed in Table 2. Brazil, despite its population of almost 200 million and its vast territory, has only 8 million people living without electricity, and China has only 11 million living without electricity even though it has seven times the population of Brazil.

Both the Chinese and Brazilian governments have overseen significant electrification efforts. For example, Brazil's rural electrification program, *Luz para Todos,* is on track to connect 10 million individuals through approximately $7 billion in grid expansion, at a cost of $3,500 per connection. Brazil relies on large hydroelectric power stations to provide more than 70 percent of the country's electricity, and it currently has excess generating capacity. As a result, its expansion has been relatively inexpensive (Niez 2010). Except in the most remote locations, electrification requires only extension of and connection to the integrated power grid.

China, in contrast, spent approximately $50 billion on its recent electricity grid development in addition to unspecified loans through state-controlled banks. This connected more than 4 million households at a higher cost per connection than Brazil and also improved quality for many millions more. In order to provide additional people with electricity, China needed to develop both additional power generation—primarily coal and distributed small hydroelectric generation—as well

providing cheap or free electricity in low-income areas may inhibit company incentives to upgrade quality. Fisher-Vanden, Mansur, and Wang (2012) analyze how electricity quality affects industry in China.

as build the corresponding electrical grid. On the other hand, the long history of electrification efforts (since at least 1949) has left most in China with some access to power, although the grid has been unreliable, unsafe, and inefficient. Indeed, as access to electricity is nearly universal, China's recent efforts have focused on quality improvements (Niez 2010). It is worth noting that Elvidge et al. (2011) use satellite images of night lighting to estimate that only about 80 percent of Chinese households use lights that are visible outdoors. They posit that one possibility for the discrepancy between their finding and official statistics is overestimation of the official electrification rate. However, on-the-ground studies (for example, UN-Energy, 2003) estimate higher electrification rates in exactly the regions where Elvidge et al. find the least electrification. It is possible that satellite images do not pick up light from the poor, who have low electricity quality and light sources that are limited and dim.

Cross-Country Correlates with Electrification

Contrast China's high electrification rate of nearly 100 percent with India's comparatively low rate of about 65 percent. China does have higher per capita GDP than India, but this does not explain the difference. In fact, among the countries with close to complete electrification, China has nearly the lowest GDP per capita. One conjecture is that the strong authoritarian government in China has facilitated infrastructure roll-out, whereas more democratic India has been less successful devoting resources to electrification. Indeed, the International Energy Administration notes that the success of both Brazil and China's electrification programs result from strong political will and sufficient funding (Niez 2010).

To explore the patterns across the developing world, Table 3 describes the relationship between electrification rates and a series of variables. Each row reports results from a separate linear regression that we estimated on a set of developing countries. The dependent variable is the country's electrification rate.[6] In each regression, per capita GDP is one explanatory variable, and then in all rows except the first one, we add a second explanatory variable. While we certainly will not interpret the coefficients in these simple cross-sectional regressions as causal, they help identify factors consistently correlated with electrification across countries.

Table 3 reflects several interesting relationships. As expected, per capita GDP is highly correlated with electrification rates in all of the specifications. Contrary to what the China–India comparison might suggest, however, governance does not appear correlated with electrification rates: neither the coefficient on a government's Polity rating (a measure of governance ranging from hereditary monarchy at one end to consolidated democracy at the other end) nor the corruption measure from World Bank (using survey data) are statistically significant. Other factors one might have thought important are also insignificant, including natural resource endowments and urbanization. Financial openness has the opposite sign from what

[6] Results are very similar if we use a log-odds ratio as the dependent variable or if we estimate each specification on the set of 44 countries for which we have all 10 explanatory variables.

Table 3
Patterns in Electrification Rates

	Log per capita GDP		*Additional explanatory variable*		*N*
Baseline	22.887***	[1.806]			81
Polity[a]	22.930***	[1.731]	0.296	[0.372]	81
Corruption[a]	24.737***	[2.529]	–4.795	[4.555]	81
Annual Growth in GDP	23.064***	[2.024]	–7.803	[36.721]	78
Gini Coefficient[a]	25.908***	[3.162]	–0.044	[0.324]	64
Poverty Gap[a]	15.540***	[4.280]	–0.980***	[0.220]	62
Change in Poverty Gap[b]	29.929***	[2.976]	0.429*	[0.240]	49
Financial Openness[a]	23.818***	[2.021]	–1.596**	[0.789]	75
Share of Population Living in Urban Areas[a]	19.429***	[2.695]	0.223	[0.138]	81
Share of GDP from Natural Resource Rents[a]	23.899***	[1.985]	–0.210	[0.187]	81

Sources: Electrification rates from International Energy Agency, *World Energy Outlook 2009.* Gini, Poverty Gap, Share of Population Living in Urban Areas and Share of GDP from Natural Resource Rents from the World Bank *World Development Indicators 2011.* Polity from Polity2 measure from Polity IV Project. Corruption from World Bank *Worldwide Governance Indicators 2011.* Financial Openness, per Lane and Milesi-Ferretti (2007), is the sum of external assets and liabilities divided by GDP; data is from the web appendix to that work.
Note: Every row reports coefficient estimates from a separate cross-country specification where the dependent variable is the country's electrification rate as a percentage as of 2008.
[a]Average reported level in 2000 or later.
[b]Average reported level in 1980–1994 minus average reported level in 1994 and later.

we would expect—that is, greater openness is correlated with a lower rate of electrification. However, this finding is not robust across various alternative specifications and could reflect that this measure captures countries with large natural resource rents that are not invested in infrastructure. (Eight of the ten countries with the highest financial openness have significant oil revenues, much of which they may have invested abroad.)

However, one factor that appears important is the income distribution within a country. While the coefficient on the Gini coefficient is not statistically significant, the coefficient on poverty gap is significantly negative, suggesting that the presence of many low-income people is negatively correlated with electrification. The poverty gap measures the amount of money that would be required to lift the income of households below the poverty line to the poverty line. Thus, for two countries with the same per capita GDP, the one with a higher poverty gap has lower electrification. The coefficient on the change in poverty gap is positive, suggesting that recent reductions in poverty are positively correlated with electrification. Although poverty is correlated with less electrification, causality may run in both directions: not only does a lower level of income reduce demand for electricity, but researchers like Lipscomb, Mobarak, and Barhan (2011) suggest that electrification may help reduce poverty.

Table 4
Energy Use by Sector
(share of total)

	Developing world	*Developed world*	*United States*
Residential	15%	20%	22%
Commercial	7%	16%	19%
Industrial	62%	40%	32%
Transport	15%	25%	28%

Source: U.S. Energy Information Administration, *International Energy Outlook 2011.*

Micro Foundations of Macro Energy Trends

We have focused thus far on understanding household-level energy use, but can the household-level patterns in asset adoption and electrification help explain the recent growth in energy consumption in the developing world? We suspect they do, based on both deductive and empirical insights.

As a starting point, Table 4 summarizes energy use by sector for developing and developed countries, along with some additional perspective from the U.S. economy. The residential sector accounts for over 20 percent of all energy used in the United States, but only about 15 percent in the developing world. There are several potential explanations for this pattern. One possibility, for example, is that industry in the developing world is supplying U.S. and other developed world consumers. However, the proportions of energy used by households are certainly consistent with the hypothesis that economic growth will lead to large gains in residential sector energy use as households coming out of poverty purchase energy-using assets. The share of energy consumed in the transport sector is also higher in the developed world, and personal vehicles account for a large part of this consumption. For example cars account for 48 percent of the energy used in transportation in the European Union (ADEME/IEEA 2007). Again, the share of households with vehicles is likely to rise quickly as poverty falls in low-income countries.

Still, the residential sector accounts for less than a quarter of total energy demand, even in the United States, so, on its face, it may seem implausible to attempt to explain total energy demand with residential consumption. But, while we have focused on refrigerators and cars as convenient devices to describe a consistent adoption pattern across countries, the basic principles we have identified likely apply to other durable assets. And, at some level, nearly all commercial and industrial processes are at least indirectly supplying consumer demand. Thus, as more consumers buy refrigerators and air conditioners as well as cell phones and other electronics, local industry will use more energy to produce at least part of the value-chain, and the commercial sector will grow to supply them.

A growing literature within environmental engineering measures the life-cycle energy and greenhouse gas emissions from different consumer actions. One approach is to use an input-output model to describe the industries that supply, for instance, appliance manufacturers and then look at the energy intensity of the inputs. Evidence from this literature suggests that about 10 percent of the energy used by a refrigerator over its lifetime is used in manufacturing. Since the same estimates assume that refrigerators last for 15 years, in the period when a large number of people are *acquiring* refrigerators for the first time, the energy from manufacturing the refrigerator, as well as from making the steel, plastic, and refrigerants that it contains, will represent a substantial portion of the refrigerator's energy use.

One environmental engineering paper documents the relationship between household income and life-cycle greenhouse gas emissions associated with U.S. households' consumption (Jones and Kammen 2011). While they consider greenhouse gas emissions and not energy use, their analysis indicates an S-shaped relationship between income and induced greenhouse gas emissions, meaning that at the low end of the income distribution, greenhouse gas emissions rise slowly with income, rise quickly at middle incomes and rise slowly again at high incomes. We suspect this relationship would be even more pronounced for consumers in the developing world, where the consumption bundles differ more between the poor and the wealthy, but this remains to be explored.

Empirically, our previous work demonstrates that the relationship between a country's energy demand and income varies substantially across the developing world. We have estimated the relationship between the log of per capita energy consumption (total consumption, across all sectors) and log of GDP per capita using a panel data set with 37 developing countries over a 27-year period from 1980 to 2006 (Gertler, Shelef, Wolfram, and Fuchs 2011). We include country fixed effects to control for any differences in the cost of producing energy or fixed differences in the demand for energy—for instance, those driven by weather. We also estimate year fixed effects to control for common price or technological shocks. We show that the coefficient on the log of GDP is considerably larger for countries with pro-poor growth. Our estimates suggest that energy demand in a country at the 75th percentile of pro-poor growth grows faster than per capita income, while energy demand in a country at the 25th percentile rises only about half as quickly as per capita income. Because our specifications do not control for other factors that could influence demand, we would not interpret the coefficients on log of per capita energy demand as true income elasticities. However, the results do suggest large differences in the relationship between income growth and energy use across countries, depending on which households benefit from growth.

Rethinking Macro Energy Forecasts

Unfortunately, we suspect that the forecasts of energy use in the developing world that we cited at the beginning of the paper do not account for the differences in growth along the extensive and intensive margins. As a result, they may seriously underestimate the likely near-term growth in energy demand.

What leads us to believe that, for example, Energy Information Administration forecasts might be wrong, and how might this be linked to asset acquisition by households coming out of poverty? For one, as we documented in Figure 1, they are projecting that the *linear* trend in energy demand growth will slow down in the developing world. This seems implausible given the large number of people in the developing world who have yet to acquire even the most basic energy-using assets (Table 1) and given the expected population growth in the developing world.

More systematically, it appears that the Energy Information Administration's past forecasts have consistently underestimated energy demand in regions that have experienced pro-poor growth. We have compared their *2000 International Energy Outlook* forecast (EIA 2000) for each region's energy demand in 2005 with actual energy demand in 2005. Adjusting for errors in forecast GDP growth, China and Brazil, had respectively approximately 15 and 10 percent higher energy consumption in 2005 than the Energy Information Administration had predicted in the 2000 report.

Both Brazil and China had pro-poor economic growth over this period. Brazil, for instance, launched the large and aggressive conditional cash transfer program, *Bolsa Família*, in 2003. The program, which currently benefits over 12 million households, has been credited with lifting 20 million people out of poverty (World Bank 2010). While China has not had an explicit anti-poverty program, it has had notable success reducing poverty since the early 1980s (for example, Ravallion and Chen 2007). In contrast, the Energy Information Administration overestimated the growth in energy demand in the rest of Central and South America, where the share of people living in poverty has recently increased (World Bank 2011a).

The forecasting model of the Energy Information Administration is complex, but thoroughly documented. The most recent documentation available suggests that the model does not fully account for potential differences in acquisition of home appliances and other assets across developing countries. For example, U.S. Energy Information Administration (2010b) reports the specific income elasticities used as inputs to the residential energy demand component of the model. The elasticity estimates are slightly higher in the developing world than the developed, but there are no differences across the developing world. Our hypothesis suggests that countries where many households are coming out of poverty (like Brazil in recent years) should have much higher income elasticities of demand for energy than countries where growth favors households at the higher end of the income distribution (like most of the rest of South America in recent years).

There are several different approaches to forecasting energy demand, an endeavor people have been engaged in for decades. The Energy Information Administration uses what would be characterized as a "top-down" model, which relies on estimates of GDP growth and energy income elasticities. In the late 1970s, in the face of declining demand, which they suspected was related to saturation of appliance holdings, the adoption of natural gas heating, and energy efficiency standards, California policymakers pioneered the application of end-use modeling, which describes residential-sector electricity demand as the sum of demand across different uses (Blumstein et al. 1979). While it might make sense for the Energy Information Administration to make country-by-country adjustments to income elasticities, it will also be valuable to develop end-use models that reflect rapid increase in the middle class within the developing world. For example, McNeil and Letschert (2010) model residential appliance uptake, although their specification is based on cross-country S-curves and does not reflect income distribution within a country.

Conclusions

Energy growth along the extensive margin, as low-income households buy their first durable appliances and vehicles, will be an important driver of the demand for energy in the near future. Within a country, the adoption of energy-using assets typically follows an S-shaped pattern: among the very poor we see little increase in the number of households owning refrigerators, vehicles, air conditioners, and other assets as incomes go up; above a first threshold income level, we see rapid increases of ownership with income; and above a second threshold, increases in ownership level off. A large share of the world's population has yet to go through the first transition suggesting there is likely to be a large increase in the demand for energy in the coming years.

Current energy forecasts appear to understate the degree to which the distribution of economic growth affects demand. These forecasts have implications for the appropriate scale of investment in the energy infrastructure. Underestimates of demand may lead to underinvestment in energy production, implying shortages and price spikes. Energy demand also has important implications for understanding the likely path of pollution, including both local pollutants and greenhouse gas emissions.

It is important to remember, however, that increases in the demand for energy associated with poverty reduction result from increases in household welfare. With a refrigerator, people may spend less time walking to stores or less time cooking. Refrigeration may affect nutrition patterns and improve health outcomes. Similarly, the switch from burning wood to using electric stoves for cooking may not only improve indoor air quality, but reduce greenhouse gas emissions because solid-fuel stoves are inefficient and gathering wood for cooking can lead to deforestation (Bruce, Rehfuess, and Smith 2011). While there is little direct evidence on the consequences of energy-using asset accumulation, Dinkelman (2011) cleverly uses plausibly exogenous variation in the cost of laying electricity distribution lines in South Africa to show that village-level electrification leads to increased female labor force participation. Lipscomb, Mobarak, and Barham (2011) use an engineering model of hydroelectric dam placement to predict county-by-county electrification in Brazil and find wide-ranging benefits of electrification, from increased employment and income to poverty reduction. Understanding the mechanisms underlying these development trends and expanding the set of outcomes analyzed remain important areas for future research.

The growth in energy demand along the extensive margin will also create some intriguing opportunities for energy policy. First, an obvious point nonetheless worth stating is that poverty reduction is unambiguously good, and keeping families in poverty is not a way to reduce energy demand. Second, to avoid shortages, price increases, and unexpected environmental impacts, each country needs to account for how poverty reduction and economic growth are likely to shape future demand for energy, and make informed investments in energy infrastructure. Third, the pervasive governmental subsidies of energy prices in the developing world are not sending the right signals for taking energy conservation or environmental externalities into account. Moreover, there is evidence from high-income countries that even if households face appropriate prices, they may make decisions about energy-using goods that are myopic. Finally, there is a chance to improve the energy efficiency of assets purchased by the large numbers of households about to come out of poverty, through energy efficiency standards, subsidized distribution of efficient and environmentally friendly models, subsidized research on energy efficient technologies, and other market interventions. Such measures could be very important, given the long lifetimes of many energy-using durable goods.

■ *We are grateful to Moshe Barach and Walter Graf for superb research assistance, to Hunt Allcott, Max Auffhammer, Randy Chugh, Lucas Davis, Taryn Dinkelman, Michael McNeil, Shaun McRae, Mushfiq Mobarak, and Omar Romero-Hernandez for useful comments and suggestions, and to Anjini Kochar, Alex Rodriguez, Jean Roth, and Terry Sicular for assistance accessing data. We would also like to thank the JEP editors.*

References

ADEME/IEEA. 2007. "Evaluation of Energy Efficiency in the EU-15: Indicators and Measures." http://www.odyssee-indicators.org/publications/PDF/brochure.pdf.

Aghion, Philippe, and Patrick Bolton. 1997. "A Theory of Trickle-Development." *Review of Economic Studies* 64(2): 151–72.

Aldy, Joseph E. 2006. "Per Capita Carbon Dioxide Emissions: Convergence or Divergence?" *Environmental and Resource Economics* 33(4): 533–55.

Auffhammer, Max. 2011. "The Relationship between Air Conditioning Adoption and Temperature." Unpublished paper, UC Berkeley, Department of Agriculture and Resource Economics.

Banerjee, Abhijit. 2004. "Contracting Constraints, Credit Markets, and Economic Development." In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society,* vol. 3, edited by Mathias Dewatripoint, Lars P. Hansen, and Stephen J. Turnovsky, 1–46. Cambridge University Press.

Banerjee, Abhijit, and Andrew F. Newman. 1993. "Occupational Choice and the Process of Development." *Journal of Political Economy* 101(2): 274–98.

Barassi, Marco R., Matthew A. Cole, and Robert J. R. Elliott. 2011. "The Stochastic Convergence of CO_2 Emissions: A Long Memory Approach." *Environmental and Resource Economics* 49(3): 367–85.

Blumstein, Carl, M. R. Jaske, T. Gorin, D. Carner, J. Carroll, S. George, B. Greene, L. Larson, D. Paige, and S. Sarvate. 1979. "Residential Electricity Demand in California: Results and Methodology." In *Changing Energy Use Futures: Proceedings of the Second International Conference on Energy Use Management,* edited by Rocco A. Fazzolare and Craig B. Smith. New York: Pergamon Press.

Bruce N. G., E. A. Rehfuess, and K. R. Smith. 2011. "Household Energy Solutions in Developing Countries." In *Encyclopedia of Environmental Health,* Vol. 3: H–M, edited by J. O. Nriagu et al., pp. 62–75. Elsevier.

Chugh, Randy. 2011. "Dieselization of the Indian Car Market, an Empirical IO Perspective."

Dargay, Joyce, and Dermot Gately. 1999. "Income's Effect on Car and Vehicle Ownership, Worldwide: 1960–2015." In *Transportation Research, Part A* 33: 101–138.

Dargay, Joyce, Dermot Gately, and Martin Sommer. 2007. "Vehicle Ownership and Income Growth, Worldwide: 1960–2030." *Energy Journal* 28(4): 143–170.

Deaton, Angus. 1986. "Panel Data from a Time Series of Cross-Sections." *Journal of Econometrics* 30(1–2): 109–126.

de Mel, Suresh, David McKenzie, and Christopher Woodruff. 2008. "Returns to Capital in Microenterprises: Evidence from a Field Experiment." *Quarterly Journal of Economics* 123(4): 1329–72.

Dinkelman, Taryn. 2011. "The Effects of Rural Electrification on Employment: New Evidence from South Africa." *American Economic Review* 101(7): 3078–3108.

Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica* 52(2): 345–62.

EIA (See U.S. Energy Information Administration)

Elvidge, Christopher D., Kimberly E. Baugh, Paul C. Sutton, Budhendra Bhaduri, Benjamin T. Tuttle, Tilotamma Ghosh, Daniel Ziskin, and Edward H. Erwin. 2011. "Who's in the Dark—Satellite Based Estimates of Electrification Rates." Chap. 15 in *Urban Remote Sensing: Monitoring, Synthesis and Modeling in the Urban Environment,* edited by Xiaojun Yang. Oxford: John Wiley & Sons.

Farrell, M. J. 1954. "The Demand for Motor-Cars in the United States." *Journal of the Royal Statistical Society.* Series A (General) 117(2): 171–201.

Fisher-Vanden, Karen, Erin T. Mansur, and Qiong Wang. 2012. "Costly Blackouts? Measuring Productivity and Environmental Effects of Electricity Shortages." NBER Working Paper 17741.

Gertler, Paul, Orie Shelef, Catherine Wolfram, and Alan Fuchs. 2011. "Poverty, Growth, and the Demand for Energy." Energy Institute at Haas Working Paper 224.

International Energy Agency. 2009. *World Energy Outlook, 2009.* Paris, France: International Energy Agency.

International Energy Agency. 2010. *World Energy Outlook, 2010.* Paris, France: International Energy Agency.

International Energy Agency. 2011. *Energy Balances of Non-OECD Countries,* 2011 Edition. Paris, France: International Energy Agency.

Jones, Christopher M., and Daniel M. Kammen. 2011. "Quantifying Carbon Footprint Reduction Opportunities for U.S. Households and Communities." *Environmental Science and Technology* 45(9): 4088–95.

Lane, Philip R., and Gian Maria Milesi-Ferretti. 2007. "The External Wealth of Nations Mark II: Revised and Extended Estimates of Foreign Assets and Liabilities, 1970–2004." *Journal of International Economics* 73(2): 223–50.

Lindh, Thomas, and Henry Ohlsson. 1998. "Self-Employment and Wealth Inequality." *Review of Income and Wealth* 44(1): 25–42.

Lipscomb, Molly, Ahmed Mushfiq Mobarak, and Tania Barham. 2011. "Development Effects of Electrification: Evidence from the Geologic Placement of Hydropower Plants in Brazil." http://faculty.som.yale.edu/mushfiqmobarak/development%20effects%20of%20electrification.pdf.

Lloyd-Ellis, Huw, and Dan Bernhardt. 2000. "Enterprise, Inequality and Economic Development." *Review of Economic Studies* 67(1): 147–68.

McNeil, Michael, and Virginia Letschert. 2010. "Modeling Diffusion of Electrical Appliances in the Residential Sector." *Energy and Buildings,* 42(6): 783–90.

McRae, Shaun. 2009. "Infrastructure Quality and the Subsidy Trap." Stanford Institute for Economic Policy Research, Discussion Paper 09-017.

National Bureau of Statistics. 2010. *China Energy Statistical Yearbook.* Beijing, China.

Ni, Chunchun. 2009. *China Energy Primer.* Lawrence Berkeley National Laboratory, LBNL-2860E.

Niez, Alexandra. 2010. *Comparative Study on Rural Electrification Policies in Emerging Economies:*

Keys to Successful Policies. International Energy Agency.

Ravallion, Martin, and Shaohua Chen. 2007. "China's (Uneven) Progress against Poverty." *Journal of Development Economics* 82(1): 1–42.

Strazicich, Mark C., and John A. List. 2003. "Are CO_2 Emission Levels Converging among Industrial Countries?" *Environmental and Resource Economics* 24(3): 263–71.

UN-Energy. 2003. "Assessing Policy Options for Increasing the Use of Renewable Energy for Sustainable Development: Modelling Energy Scenarios For Sichuan, China." A UN-Energy Demonstration Study. Conducted by Department of Economic and Social Affairs (DESA), International Atomic Energy Agency (IAEA), and United Nations Environmental Programme (UNEP). http://www.iaea.org/OurWork/ST/NE/Pess/assets/sichuan_case_study.pdf.

U.S. Department of Energy. 2011. *2010 Buildings Energy Data Book*.

U.S. Energy Information Administration (EIA). 2000. *International Energy Outlook 2000*. Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2008. *International Energy Annual 2006*. Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2010a. *International Energy Outlook 2010*. Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2010b. *World Energy Projection System Plus: Residential Model*. Washington, DC: EIA.

U.S. Energy Information Administration (EIA). 2011. *International Energy Outlook 2011*. Washington, DC: EIA.

World Bank. 2010. "Lifting Families Out of Poverty in Brazil—Bolsa Familia Program." Webpage, updated September 27. http://go.worldbank.org/QCZI04L470.

World Bank. 2011a. *World Development Indicators*.

World Bank. 2011b. *Worldwide Governance Indicators*.

Zhou, Nan, David Fridely, Michael McNeil, Nina Zheng, Virginie Letschert, Jing Ke, and Yarmina Saheb. 2011. "Analysis of Potential Energy Saving and CO_2 Emission Reduction of Home Appliances and Commercial Equipments in China." *Energy Policy* 39(8): 4541–50.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 139–164

The For-Profit Postsecondary School Sector: Nimble Critters or Agile Predators?†

David J. Deming, Claudia Goldin, and Lawrence F. Katz

Private for-profit institutions have become an increasingly visible part of the U.S. higher education sector. Within that sector, they are today the most diverse institutions by program and size, have been the fastest growing, have the highest fraction of nontraditional students, and obtain the greatest proportion of their total revenue from federal student aid (loan and grant) programs. They are, as well, the subjects of high-profile investigations and are facing major regulatory changes.

Today's for-profit postsecondary schools were preceded a century ago by a group of proprietary schools that were also responding to an explosion in demand for technical, vocational, and applied subjects. Business, managerial, and secretarial skills were in great demand in the late nineteenth and early twentieth centuries, and a multitude of proprietary institutions emerged that taught accounting, management, real estate, stenography, and typing. The numbers and enrollments of these institutions were greatly reduced when public high schools expanded and increased their offerings in the business and vocational areas. But many survived and morphed into some of the current for-profits, such as Blair College (established 1897; now part of Everest College), Bryant and Stratton College (1854),

■ *David J. Deming is Assistant Professor of Education and Economics, Graduate School of Education, Claudia Goldin is the Henry Lee Professor of Economics, and Lawrence F. Katz is the Elisabeth Allison Professor of Economics, all at Harvard University, Cambridge, Massachusetts. Goldin is Director of the Development of the American Economy program and Katz is a Research Associate at the National Bureau of Economic Research, Cambridge, Massachusetts. Their e-mail addresses are ⟨david_deming@gse.harvard.edu⟩, ⟨cgoldin@harvard.edu⟩, and ⟨lkatz@harvard.edu⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.1.139.

doi=10.1257/jep.26.1.139

Gibbs College (1911), Globe University (1885), Rasmussen College (1900), and Strayer University (1892).

Distance learning, known today as online education, also has an interesting past in "correspondence courses" that were offered by many universities beginning in the late nineteenth century including some of the most prestigious, such as the University of Chicago and the University of Wisconsin (Watkins 1991). Online education is today's most rapidly growing part of higher education. Walden University, founded in 1970 and today one the largest for-profit online institutions, pioneered online studies to allow working professionals to earn further degrees.

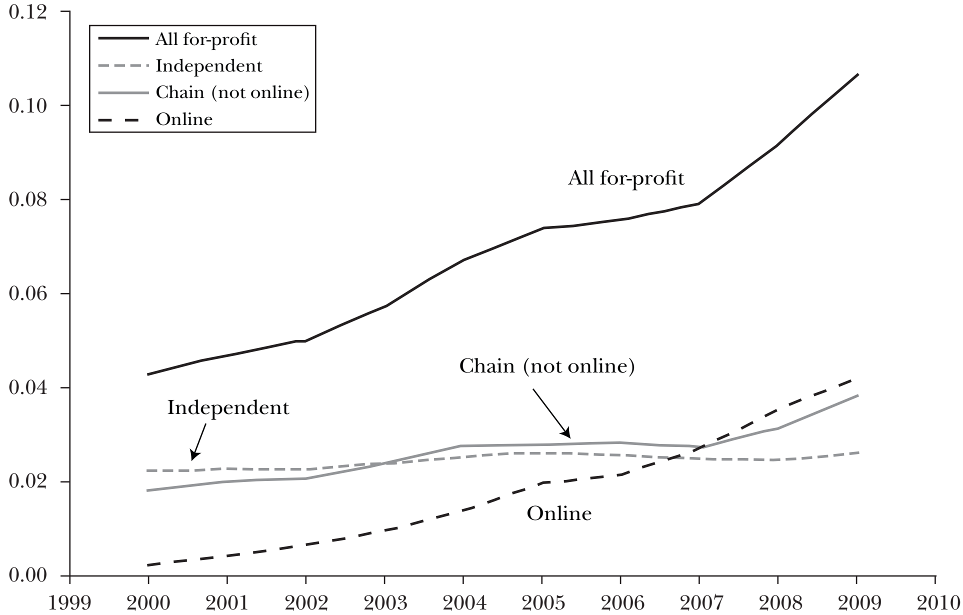
In this article, we describe the schools, students, and programs in the for-profit higher education sector, its phenomenal recent growth, and its relationship to the federal and state governments. As a starting point, for-profit postsecondary enrollments have grown considerably during the past several decades, particularly in degree programs and at large national providers with substantial online offerings. Fall enrollment in for-profit degree-granting institutions grew by more than 100-fold from 18,333 in 1970 to 1.85 million in 2009. During that same time period, total fall enrollment in all degree-granting institutions increased 2.4-fold from 8.58 million in 1970 to 20.43 million in 2009 (U.S. Department of Education, NECS, 2010, *Digest*, table 197). Thus, for-profit enrollment increased from 0.2 percent to 9.1 percent of total enrollment in degree-granting schools from 1970 to 2009. For-profit institutions for many decades also have accounted for the vast majority of enrollments in non-degree-granting postsecondary schools (those offering shorter certificate programs), both overall and among such schools eligible for federal (Title IV) student financial aid.

Figure 1 highlights the rise of for-profits in the enrollments of Title IV–eligible (degree and non-degree-granting) higher education institutions since 2000, a period when enrollment in the for-profit sector tripled while enrollment for the rest of higher education increased by just 22 percent. The solid dark line shows that the fraction of fall enrollments accounted for by the for-profits increased from 4.3 percent in 2000 to 10.7 percent in 2009. For the descriptive data presented here, we rely extensively on the Integrated Postsecondary Education Data System (IPEDS) of the U.S. Department of Education, which is an annual survey of all postsecondary institutions that participate in the federal student financial aid programs.[1]

Under the solid dark line in Figure 1, the growth of the for-profit sector is broken down into "independent" schools, online institutions, and for-profit "chains." We must first define these terms, because these categories are not designated in the official IPEDS data. "Independent" schools are defined here as

[1] An online Appendix available with this paper at ⟨http://e-jep.org⟩ provides the details of our processing of the micro IPEDS data, linkage of the IPEDS institution-year data to financial aid to data from the National Student Loan Data System, and construction of an institution-level panel data set for 2000 to 2009.

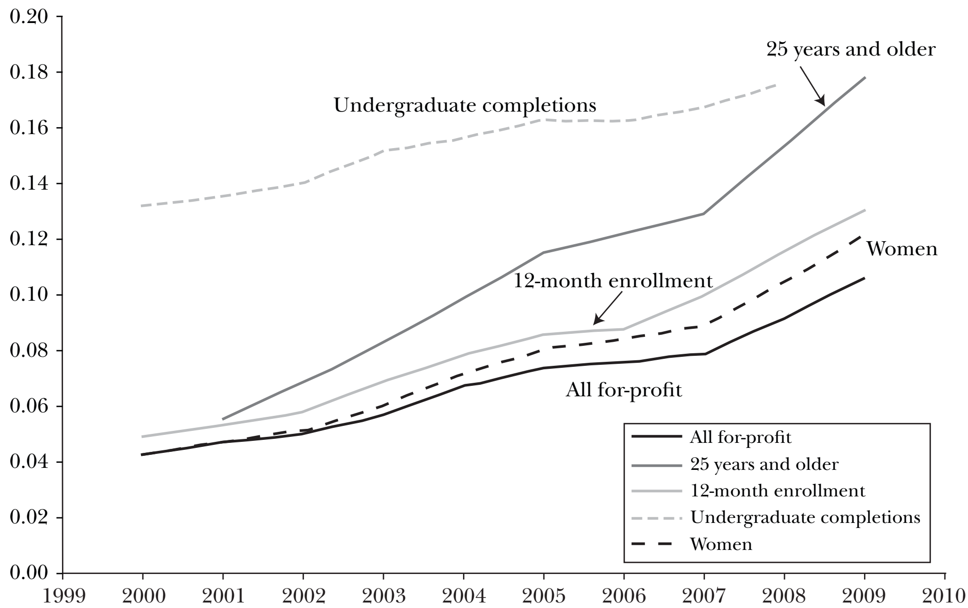
Figure 1
For-Profit Institution Share of Total Title IV Fall Enrollment: Total and by School Type, 2000 to 2009



Source: Integrated Postsecondary Education Data System (IPEDS).
Notes: A for-profit institution is classified as "online" if it has the word online in its name or if not more than 33 percent of the school's students are from one U.S. state. The "chain (not-online)" category covers all other for-profit institutions that operate in more than one state or have more than five campus branches within a single state. The "independent" category includes for-profits that operate in only one state and have fewer than five campus branches. An online Appendix available with this paper at ⟨http://e-jep.org⟩ provides the details of our processing of the micro IPEDS institution-level data for 2000 to 2009.

those operating in no more than one state and having no more than five campus branches. A "chain" is a for-profit institution that operates in more than one state or has more than five campus branches within a single state. A for-profit is designated as online if it has the word "online" in its name or, more commonly, if no more than 33 percent of the school's students are from one U.S. state. All online institutions are considered to be chains because they serve students in multiple geographic markets. Independent schools showed little increase in their share of overall enrollments in higher education from 2000 to 2009; chains with largely in-person enrollment showed a doubling over this period; and online institutions, typically part of national publicly traded companies, increased from almost nothing to become the largest part of the sector. Indeed, almost 90 percent of the increase in for-profit enrollments during the last decade occurred because of the expansion of for-profit chains.

Figure 2
For-Profit Share of Enrollments and Undergraduate Completions: 2000 to 2009



Source: Integrated Postsecondary Education Data System (IPEDS).
Notes: "All for-profit" is fall enrollment, that is enrollment at the beginning of the academic year; "12-month enrollment" = unduplicated enrollment during the entire year; "25 years and older" = fall enrollment of those 25 years and older; "women" = female fall enrollment; "undergraduate completions" = all undergraduate completions (certificates + associate's degrees + bachelor's degrees). The series for "25 years and older" is for the odd-numbered years and the even-numbered years are interpolated from those.

The rapid growth of the for-profits from 2000 to 2009 is illustrated in various ways in Figure 2. The for-profit share of enrollments (unduplicated headcount) over a 12-month period increased from 5 percent in 2001 to 13 percent in 2009. The 12-month enrollment measure better captures enrollments in for-profits than the standard fall enrollment measure because it includes students in less-conventional and short programs entered throughout the year.

For-profits have expanded their enrollment share more rapidly for women than for men, and they play an increasingly large role in the higher education of older students. The for-profit enrollment share of students 25 years and older expanded from around 6 percent in 2001 to 18 percent in 2009. Undergraduate completions from for-profit institutions grew from 13 percent of the total in 2000 to almost 18 percent in 2008. The fraction of completions is considerably larger than that for enrollments because more than half of for-profit completions are certificates and most certificate programs are no more than one year.

For-profit enrollments and completions in recent years have been growing most rapidly in longer degree programs. In the last decade, the for-profits increased their share of completers in all types of undergraduate programs, but more so for

AAs (associates' degrees) and BAs (bachelor's degrees) than for certificates. They produced about 39 percent of certificates in 2000 and 42 percent in 2008. For-profit AAs were 13 percent of all AAs in 2000 but 18 percent in 2008; BAs were less than 2 percent of all in 2000 but were 5 percent of all BAs in 2008 (U.S. Department of Education, NECS, 2010, *Digest*, table 195).

The current incarnation of the for-profit sector is big business; its largest providers are major, profitable, publicly traded corporations (Bennett, Lucchesi, and Vedder 2010). They appear to be nimble critters that train nontraditional learners for jobs in fast-growing areas, such as health care and information technology. On the other side, most of them depend on U.S. government student aid for the vast bulk of their revenues. Default rates on the loans taken out by their students vastly exceed those of other institutions of higher education, and audit studies have shown that some for-profits have engaged in highly aggressive and even borderline fraudulent recruiting techniques (U.S. Government Accountability Office 2010).

Are the for-profits nimble critters or agile predators? Using the 2004 to 2009 Beginning Postsecondary Students (BPS) Longitudinal Study, we assess outcomes of a recent cohort of first-time undergraduates who attended for-profits relative to comparable students who attended community colleges or other public or private nonprofit institutions. We find that relative to community colleges and other public and private nonprofits, for-profits educate a larger fraction of minority, disadvantaged, and older students, and they have greater success at retaining students in their first year and getting them to complete shorter degree and nondegree programs at the certificate and AA levels. But we also find that for-profits leave students with far larger student loan debt burdens. For-profit students end up with higher unemployment and "idleness" rates and lower earnings from employment six years after entering programs than do comparable students from other schools. Not surprisingly, for-profit students have trouble paying off their student loans and have far greater default rates. And for-profit students self-report lower satisfaction with their courses of study and are less likely to consider their education and loans worth the price-tag relative to similarly-situated students who went to public and private nonprofit institutions.

What is the For-Profit Postsecondary School Sector?

Apollo and the Lesser For-Profit Deities: A Diverse Sector

The for-profit postsecondary school sector, at its simplest level, is a group of institutions that give post-high school degrees or credentials and for which some of the legal "nondistribution requirements" that potentially constrain private nonprofit schools do not bind. For example, for-profit institutions can enter the equity market and have few constraints on the amounts they can legally pay their top managers. In practice, only the largest players in this market raise substantial capital in organized equity markets, and they tend to pay their top executives mega-salaries that exceed those of presidents at the public and nonprofit private universities.

Among the for-profits, Andrew Clark, chief executive officer of Bridgepoint Education, Inc., received more than $20 million in 2009, while Charles Edelstein, co-chief executive officer of the Apollo Group, Inc., earned more than $11 million.[2]

For-profit sector institutions are a varied group. For-profit schools offer doctorates but also nondegree courses, and their programs run the gamut from health care, business, and computers to cosmetology, massage, and dog grooming. The sector contains the largest schools by enrollment in the United States and also some of the smallest. For example, the University of Phoenix Online campus enrolled over 532,000 students, and Kaplan University enrolled 96,000 during the 2008–2009 academic year. Taken together, the largest 15 institutions account for almost 60 percent of for-profit enrollments (Bennett, Lucchesi, and Vedder 2010, table 1). But tabulations from the IPEDS also indicate that the median Title IV–eligible, for-profit institution had a Fall 2008 enrollment of 172 students as compared with 3,713 for the median community college (two-year public institution), 7,145 for the median four-year public university, and 1,149 for the median four-year, private not-for-profit school.

The for-profit sector has become in many people's minds synonymous with the large for-profit chains that have rapidly expanded their presence in the BA and graduate education markets, especially the Apollo Group, which owns the University of Phoenix. But even though the big players in this sector do account for the majority of for-profit enrollments, another important part of the sector consists of career colleges that focus on a wide range of shorter AA and certificate programs. Completions in the for-profit sector are still dominated by certificate programs, and 55 percent of the certificates granted by the for-profits are awarded by the 1,700 or so independent career colleges and institutes. Our tabulations from the IPEDS indicate that certificates account for 54 percent of the degrees and awards conferred by for-profits in 2008–2009.

There are several important commonalities across this mixed group. The for-profit sector offers almost no general education and liberal arts programs. For-profit programs typically are not meant to prepare students to continue to another form of higher education, as is the case with most community colleges. Rather, the for-profits almost always offer training for a vocation or trade. In that sense, they are "career colleges." In addition, virtually all the for-profits require that admitted students have a high school diploma or another secondary school credential such as a GED. Their ability to obtain federal (Title IV) financial aid for their students is typically contingent on their admitting primarily students who have already

[2] In higher education, nonprofits and publics are not that far behind in pay, just below the very top of the for-profit scale. In 2006/07, before the stock market decline, the highest paid university president was Gordon Gee at Vanderbilt who earned slightly more than $2 million in total compensation. A bit lower down the scale, the tenth highest-paid CEO at a for-profit was Wallace Boston, Jr., CEO of American Public Education, with $961,000, while number 10 among the presidents of public institutions on the list was Jack Varsalona at Wilmington University who earned $974,000. After the stock market drop, earnings in 2008/09 for presidents at public and nonprofit private universities were far lower. The data on for-profit CEO pay is from *Chronicle of Higher Education* (2010); data on public and nonprofit president's pay is from Gibson (2009).

completed secondary school. However, beyond requiring a high school degree, for-profit institutions are almost always nonselective and open admissions.

For-profit higher education is more likely to flourish in providing vocational programs that lead to certification and early job placement—programs that have clear short-run outcomes that can serve to build institutional reputation in the labor market. But the for-profits are likely to be in a far less advantageous position where external benefits (and subsidies from donors and government) are important and where the qualities of inputs and outputs are difficult to verify (Winston 1999). For-profits also have been successful at designing programs to attract nontraditional students who may not be well-served by public institutions (Breneman, Pusser, and Turner 2006).

What is Title IV Eligibility?

The for-profit sector that we analyze here includes almost exclusively those that are termed "Title IV eligible." Because for-profits often cater to independent students and those from low-income families who finance college through Pell grants and federal student loans, they have an intricate relationship with the federal government to ensure they maintain eligibility to receive Title IV federal student aid. The for-profits, like public institutions of higher education, receive an extremely large fraction of their revenues from government sources.

Title IV eligibility is granted by the U.S. Department of Education and requires that the institution be accredited by at least one of their approved accrediting agencies, be registered by one of the states, and meet other standards on a continued basis. Some of these standards concern the length of programs and some concern students and their federal loan repayment activity. A Title IV–eligible, private for-profit school must either provide training for gainful employment in a recognized occupation or provide a program leading to a baccalaureate degree in the liberal arts (U.S. Department of Education 2011a). Our discussion excludes non–Title IV, for-profit schools, about which little has been known because the U.S. Department of Education does not track them. Virtually all degrees are granted by Title IV–eligible institutions, but programs that are less than two years in length that grant certificates (also diplomas) often are found at non–Title IV institutions. For an analysis of the importance of the non–Title IV group of for-profit schools using state registration data, see Cellini and Goldin (forthcoming). Because virtually all degree-granting institutions are Title IV–eligible, the undercount from limiting the analysis to Title IV schools impacts only the nondegree (typically certificate) programs in institutions without any degree program.

For-Profit Programs

The for-profits loom large in the production of degrees and certificates in certain programs. For-profits produce 18 percent of all associate's degrees, but they produce 33 percent of the AAs granted in business, management, and marketing, 51 percent in computer and information sciences, 23 percent in the health professions, and 34 percent in security and protective services. In the public and nonprofit private sectors, an AA degree is often the gateway to a four-year college and, in consequence, 38 percent of these AA programs are in general studies and liberal arts programs. In the for-profits, a mere 2.4 percent are in general studies and liberal arts.

Although 5 percent of all BAs are granted by for-profit institutions, 12 percent of all BAs in business, management, and marketing are. Other large for-profit BA programs are in communications (52 percent of all BAs in communications are granted by for-profits), computer and information sciences (27 percent), and personal and culinary services (42 percent).

Certain programs are highly concentrated in the for-profit degree categories. Among AA degrees just two program groups—business, management, and marketing, and the health professions—account for 52 percent of all degrees. In the BA group, the business program produces almost 50 percent of the total. Among certificates granted in the Title IV for-profit sector, health professions and personal and culinary services account for 78 percent of certificate completers (U.S. Department of Education, NCES, 2009, tables 37 and 40; authors' tabulations from the IPEDS).

Who Are the Students?

The for-profit sector disproportionately serves older students, women, African-Americans, Hispanics, and those with low incomes. Table 1 looks at the characteristics of students in various types of institutions of higher education. African Americans account for 13 percent of all students in higher education, but they are 22 percent of those in the for-profit sector. Hispanics are 11.5 percent of all students but are 15 percent of those in the for-profit sector. Women are 65 percent of those in the for-profit sector. For-profit students are older: about 65 percent are 25 years and older, whereas just 31 percent of those at four-year public colleges are, and 40 percent of those at two-year colleges are.

Using the Beginning Postsecondary Students longitudinal survey data for students entering postsecondary school during the 2003–2004 academic year, we can get a more detailed picture of for-profit students relative to those at other colleges. Because the BPS surveys only first-time undergraduates, the results are somewhat different from the IPEDS, which surveys institutions about all students. But the storyline remains the same.

Compared with those in community colleges (almost entirely two-year public schools), for-profit students are disproportionately single parents, have much lower family incomes, and are almost twice as likely to have a General Equivalency Degree (GED). Among for-profit students in the Beginning Postsecondary Students data, 55 percent are in certificate programs and just 11 percent are enrolled in a BA program. Similarly, among all for-profit students in the IPEDS, certificates are 54 percent of all completions or degrees conferred, and associates are 22.5 percent (U.S. Department of Education, NECS 2010, *Digest,* table 195). The BA group is just 13 percent but is the fastest-growing degree group among the for-profits. Postgraduate programs, primarily master's degrees, account for the remaining 10.5 percent.[3]

[3] We should note that the comparison between enrollments in the Beginning Postsecondary Students data and completions in the IPEDS is generally not valid when programs vary in length. But because the

Table 1
Student Characteristics from the BPS and IPEDS for For-Profits, Two-Year Public Colleges, and Four-Year (Nonprofit) Colleges

	Student characteristics by IPEDS institution type, 2009/2010			
	For-profit institutions	*2-year public colleges*	*4-year public colleges*	*4-year private nonprofit colleges*
Female	0.651	0.570	0.552	0.576
African-American	0.221	0.136	0.109	0.104
Hispanic	0.150	0.157	0.105	0.093
Full-time	0.579	0.410	0.733	0.742
Age 25 years and over	0.651	0.404	0.306	0.392
Federal loans per student	11,415	759	3,512	5,769
Pell Grant per student	2,370	773	738	632
Tuition (in-state)	13,103	2,510	5,096	24,470
Number of institutions	2,995	1,595	690	1,589

	BPS 2004–2009 sample characteristics		
	For-profit institutions	*Community colleges*	*4-year public and nonprofit colleges*
Female	0.659	0.564	0.558
African-American	0.248	0.140	0.141
Hispanic	0.264	0.159	0.103
Age	24.4	23.8	19.5
Single parent	0.288	0.124	0.030
Delayed enrollment after high school	0.576	0.481	0.142
High school diploma	0.754	0.852	0.947
GED	0.172	0.095	0.022
Mother high school dropout	0.224	0.137	0.055
2003 family income if a dependent	36,854	60,039	76,509
2003 family income if independent	17,282	31,742	78,664
Enrolled full-time	0.809	0.460	0.903
Worked while enrolled, 2003–2004	0.635	0.755	0.499
Enrolled in a certificate program	0.551	0.072	0.015
Enrolled in an AA program	0.326	0.774	0.061
Enrolled in an BA program	0.106	0	0.891
Expects to earn a BA	0.643	0.799	0.980
Sample size (unweighted)	1,950	5,970	8,760

Sources: BPS:04/09, or Beginning Postsecondary Students Longitudinal Study data for 2003–2004 first-time beginning postsecondary students in their first, third, and sixth years since entering an undergraduate institution, through 2009; and Integrated Postsecondary Education Data System (IPEDS) data.
Notes: Community colleges include two-year public and private nonprofit institutions. Unweighted sample sizes in the BPS data are rounded to the nearest 10. The IPEDS tabulations cover the (undergraduate and graduate) enrollments of Title IV institutions in Fall 2009. The BPS tabulations cover beginning postsecondary students entering a Title IV institution in the 2003–2004 academic year.

BPS surveys a cohort, the comparison has greater validity.

The Business Model of the For-Profit Sector

For-profit chains led by online institutions experienced phenomenal growth in the past several decades. The growth has been largely due to an extension of a business model that has emphasized the special client base of the for-profits combined with the ability to "clone" successful programs using web technology and the standardization of curriculum for traditional in-person courses. In this section, we turn to the financial and business aspects of the for-profits. For more detail on the business strategies of for-profit colleges, the interested reader might start with Breneman, Pusser, and Turner (2006) and Hentschke (2010).

The expansion of the chains (including online institutions) accounts for 87 percent of the increase in fall enrollment during the past decade. The increase in online enrollment alone accounts for 54 percent of the total. The rise of the chains is responsible, as well, for 80 percent of the increase in federal loan and grant volumes of the for-profits. For-profit chains and online programs also benefit from economies of scale in advertising and recruitment costs.

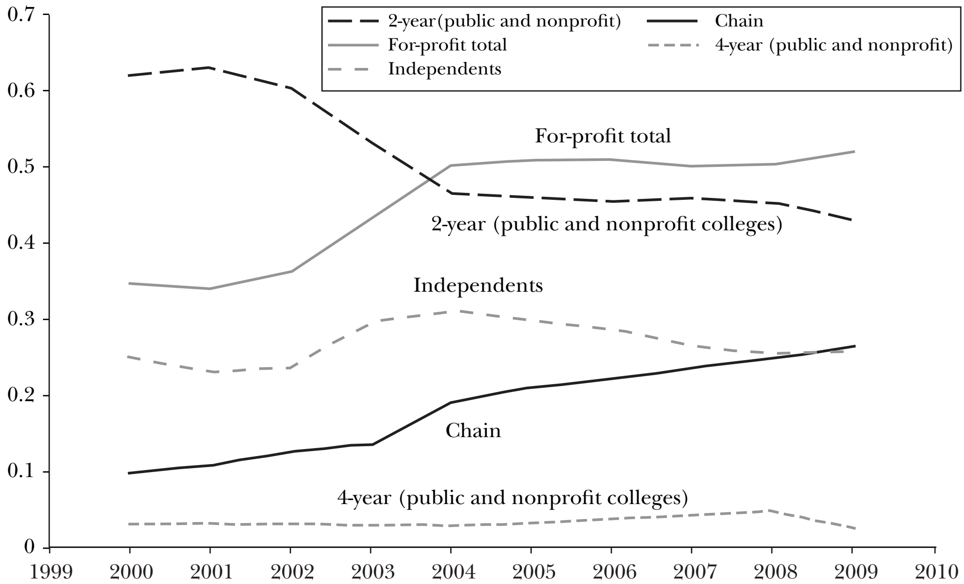
Client Base and Recruiting

The Title IV–eligible, for-profit sector receives the majority of its revenues from federal financial aid programs in the form of loans and grants to their students. For-profits appeal to older individuals who are simultaneously employed and in school or taking care of family members. Some of the for-profits offer services, such as child care, to deter enrollees from dropping out, especially during the period when the student can get a refund and to minimize the institution's dropout rate to maintain accreditation (for example, Rosenbaum, Deil-Amien, and Person 2006). The for-profits are attractive to nontraditional students, many of whom are low income, require financial aid, and need help filling out aid forms. For-profits often give generous transfer credit to students who began their BAs at other institutions.

For-profit institutions devote substantial resources to sales and marketing. Advertising in 2009, as demonstrated in one study of 13 large national chains, was around 11 percent of revenue. Sales and marketing (including advertising) for this group was around 24 percent of revenue. In consequence, the average new student recruit costs one of the large national chains about $4,000 (Steinerman, Volshteyn, and McGarrett 2011).[4] Annual tuition at for-profit institutions was about $16,000 for a BA program, $15,000 for an AA program, and $13,000 for a certificate program in 2010–11, as compared to average undergraduate tuition of about $7,000 at public four-year institutions for in-state students and $16,000 for out-of-state students, and $22,000 for private nonprofit schools (Knapp, Kelley-Reid, and Ginder 2011, table 3).

[4] The large national chains in the study are American Public Education, Apollo Group, Bridgepoint Education, Capella Education, Career Education, Corinthian Colleges, DeVry Inc., Education Management, Grand Canyon Education, ITT Educational Services, Lincoln Education, Strayer Education, and Universal Technical Institute.

Figure 3
Enrollment in Allied Health Fields by Type Institution



Source: Integrated Postsecondary Education Data System (IPEDS).
Note: "4-year (public and nonprofit colleges)" = public and private nonprofit four-year institutions; "2-year (public and nonprofit colleges)" = two year public (community colleges) and two-year private nonprofit colleges; "independents" = for-profit independent (non-chain) institutions; "chain" = for-profits institutions with "online" in the school name or that operate in more than one state or that have more than five campus branches in a single state.

Responsiveness to Markets

For-profits cater to the expanding market of nontraditional students, develop curriculum and teaching practices to be able to provide identical programs at multiple locations and at convenient times, and offer highly-structured programs to make timely completion feasible (Hentschke 2010). For-profits are attuned to the marketplace and are quick to open new schools, hire faculty, and add programs in growing fields and localities. For example, Turner (2006) finds that change in for-profit college enrollments are more positively correlated with changes in state college-age populations than are changes in public sector college enrollments.

For-profits are less encumbered than public and nonprofit schools by physical plant, alumni, and tenured faculty. Take the expanding health profession fields, for example. Enrollment in programs involving the health professions doubled from 2000 to 2009. In the for-profit sector, it tripled, whereas in all other postsecondary institutions it increased by 1.4 times. In consequence, the fraction of enrollment in the allied health fields in the for-profits increased from 35 percent to 52 percent, as illustrated in Figure 3. The increase in such enrollments at the national and regional chains accounts for almost the entire 17 percentage point increase.

Looking more closely at these programs, the for-profits have rapidly entered the growing fields of medical assisting, phlebotomy, x-ray and ultrasound technicians, practical nursing, and even registered nursing. The total number of AA degrees in the health professions doubled during the past decade, but degrees in this area from for-profits quadrupled, with degrees from the large for-profit chains rising by a multiple of six. A similar pattern arises for certificates in the health professions, where for-profit national and regional chains more than tripled their awards from 2000 to 2009 at a time when the public sector only more than doubled theirs.

Online Education

Online education fits many of the features of the for-profit business model. For example, it attracts older students who need to combine work with schooling and appeals to students who do not want to learn on the academic calendar. (There is even a popular advertisement: "Earn your college degree in your pajamas.") Much of the growth of for-profits during the last decade has been in schools emphasizing online programs, as seen in Figure 1.

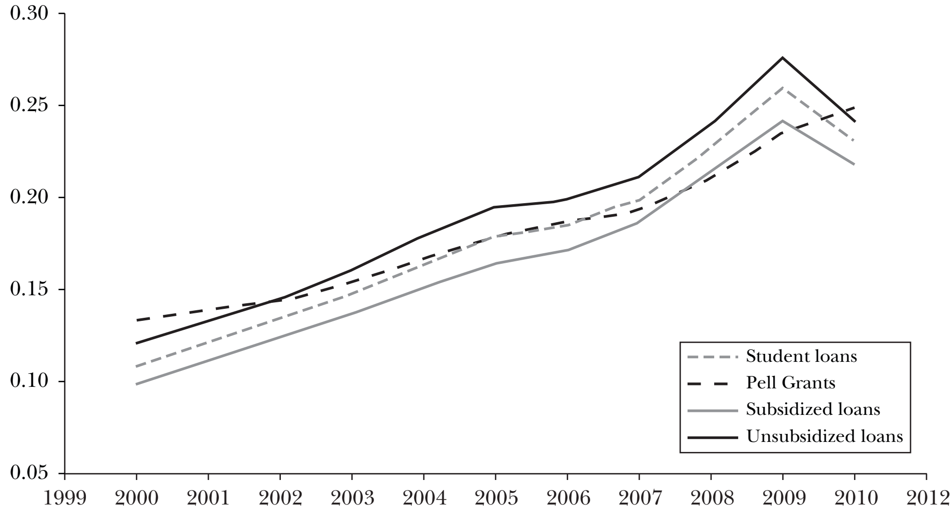
Some of this increase was due to U.S. Department of Education regulatory changes. Prior to 1998, a Title IV–eligible institution could not have more than half of its enrollment in distance education. Then in 1998, the Higher Education Act authorized the U.S. Department of Education to grant waivers to promote new advances in distance education. By the early 2000s many of the larger chains were granted waivers, and the limit on share of enrollment in distance education was dropped. The regulatory change in 2005 spurred the growth of dedicated online institutions. By 2007–2008, 12 percent of undergraduates and 25 percent of graduate students at for-profits took their entire program through distance education as compared with less than 3 percent for undergraduates and 8 percent for graduate students at public and private nonprofit institutions combined (U.S. Department of Education, NCES, 2011, tables A-43-1 and A-43-2).

Federal Student Financial Aid

Federal student financial aid is the lifeblood of for-profit higher education. Federal grants and loans received under Title IV of the Higher Education Act accounted for 73.7 percent of the revenues of Title IV–eligible, private for-profit higher education institutions in 2008-09 (based on data in U.S. Department of Education, Federal Student Aid Data Center 2011). Under current regulations, for-profit schools can derive no more than 90 percent of their revenue from Title IV financial aid sources to maintain Title IV eligibility, and the constraint comes close to binding for many for-profits. In fact, 30 percent of for-profit institutions, including many of the largest national chains such as the University of Phoenix and Kaplan University, received more than 80 percent of their revenues from federal Title IV student aid in 2008–2009. These Title IV revenue figures actually understate the importance of federal student aid to for-profit institutions since they do not include military educational benefits provided to veterans and active service members, which do not count

Figure 4

For-Profit Share of Federal Financial Aid (Pell Grants and Student Loans): 2000 to 2010



Source: National Student Loan Data System (NSLDS).
Note: Student loans include subsidized and unsubsidized federal student loans under the Federal Family Education Loan (FFEL) and Direct Loan Programs.

towards the limit of 90 percent federal Title IV student aid revenues. The for-profits have, in consequence, actively recruited military benefit recipients—veterans, service members, and their family members—especially under the Post–9/11 GI Bill of 2008. For-profits accounted for 36.5 percent of the benefits paid under the Post–9/11 GI Bill during the first year of the program (Health, Education, Labor and Pensions Committee 2010, p. 4).

For-profit institutions receive a disproportionate share of federal Title IV student financial aid both because they have higher tuition and fees than public institutions and because they attract large numbers of students who are financially independent or come from low-income families. For-profits accounted for 24 percent of Pell grant disbursements and 26 percent of federal student loan disbursements in 2008–2009 even though they enrolled 12 percent of the students (authors' tabulations from the IPEDS and NSLDS). Half of undergraduates at for-profit schools received Pell grants, as compared with 25 percent at public and private nonprofit institutions combined.

The sharp increase in the enrollments at for-profit schools has been accompanied by a rapid rise in their share of federal student financial aid from 2000 to 2010, as shown in Figure 4. The for-profit share of Pell grants increased over the last decade from 13 to 25 percent and their share of total federal student loans (both subsidized and unsubsidized loans) increased from 11 percent in 2000 to 26 percent in 2009 before dipping to 23 percent in 2010.[5]

Of course, public sector institutions receive direct taxpayer support largely from state government appropriations, enabling tuition and fees to be lower than they otherwise would be. If federal student loans to students at for-profits are repaid, taxpayer costs are actually lower to finance education in for-profits than in public sector institutions. But the comparison is not quite apples-to-apples. The rationale for subsidies to public institutions and private nonprofit schools is that they produce research with potentially large spillover benefits and that they educate students in the liberal arts and other fields that may improve civil society and generate external benefits. Also, loans to students attending for-profits often do not get repaid.

Default Rates

Students from for-profit institutions have higher default rates on federal student loans than students in other sectors. And the default rates of for-profits have risen substantially during the last five years.[6]

The two-year "cohort default rate" measures the percentage of borrowers who enter repayment of federal student loans (by leaving a program through graduation or dropping out) during a fiscal year and default prior to the end of the next fiscal year. An institution loses Title IV eligibility if its two-year cohort default rate exceeds 25 percent for three consecutive years or is 40 percent in any one year. The two-year cohort default rate of for-profit institutions was 11.6 percent for fiscal year 2008 as compared with 6 percent for public institutions and 4 percent for private nonprofits. The U.S. Department of Education is moving to a three-year cohort default rate standard for maintaining Title IV eligibility in fiscal year 2012. Three-year cohort default rates for fiscal year 2008 were 24.9 percent for for-profits, 7.6 percent for private nonprofits, and 10.8 percent for public institutions (Steinerman, Volshteyn, and McGarrett 2011). The sharp increase in default rates from a two- to a three-year window may, to some extent, reflect incentives for institutions to minimize defaults within the current two-year regulatory window. Thus, three-year default rates also are likely to provide a more realistic indicator of long-run loan repayment rates than the two-year default rates.[7]

We examine the role of student demographics, financial aid take-up, and institutional characteristics (degree types, distance education, remedial course offerings, and student services) in explaining the higher federal student loan default rates of

[5] The slight decline in the for-profit share of loans in 2010 may reflect the shift from the Federal Family Education Loan program with bank lending under federal guarantees to the Direct Loan program where the federal government makes the loans directly to students.

[6] Current default rates at for-profits, however, remain lower than in the late 1980s and early 1990s before the 1992 amendments to the Higher Education Act that tightened institutional eligibility for Title IV funds and removed many nondegree proprietary schools with very high default rates from the Title IV financial aid programs (Bennett, Lucchesi, and Vedder 2010).

[7] Furthermore, since federal Stafford loans have an initial 6-month grace period and can be up to 360 days delinquent before being considered in default, the two-year default rates typically cover a much shorter window in which a recorded default is possible.

for-profit institutions. Figure 5 graphs (regression-adjusted) differences in three-year cohort default rates by type of institution. The differences are computed from regressions of default rates on institution type (with public four-year institutions as the base group) including year dummies plus successive additions of controls for student and institution characteristics, geography, and school selectivity for pooled institution-year data covering the 2005 to 2008 fiscal years.

The raw default rates and those regression-adjusted for institutional and student characteristics are highest for the for-profit schools, followed by community colleges and then four-year public and nonprofit institutions. The unadjusted 11 percentage point higher three-year cohort default rates for for-profits (column 1) relative to the base group of four-year public institutions is reduced slightly to 10.5 percentage points with the addition of detailed controls for student demographics, institutional characteristics, and city fixed effects (columns 2 and 3) despite the fact that these controls explain a substantial fraction of the cross-institution variation in default rates. The addition of the covariates modestly expands the for-profit default rate gap relative to community colleges.

The for-profit default rate is 8.7 percentage points higher than that for four-year publics and nonprofits and 5.7 percentage points higher than for community colleges even when the sample is limited to nonselective (open admission) institutions (column 4). Higher three-year cohort default rates are apparent for all segments of the for-profit sector, including independent schools, regional chains, national chains, and largely online institutions (see Appendix Table 1, available online with this paper at ⟨http://e-jep.org⟩). National chains have higher default rates and online institutions lower default rates relative to all for-profits.

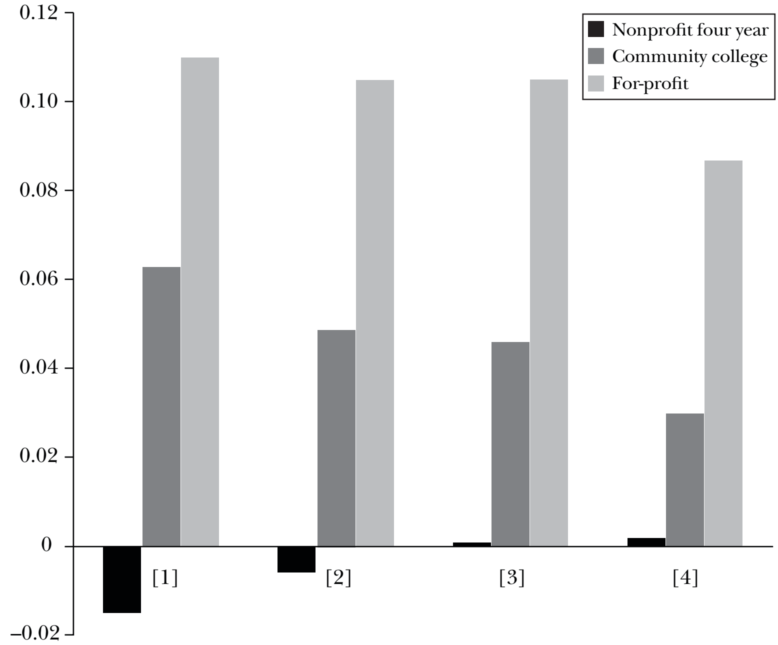
For-profit institutions account for a large and rising share of federal financial aid. For-profit students have much higher default rates than those at other schools even adjusting for differences in student characteristics. In the most recent data, they account for 47 percent of deafults. In addition, default rates have been rising particularly for the for-profit chains.

Student Outcomes

The large increase in federal student aid dollars flowing to for-profits has attracted substantial scrutiny about the quality of their programs and whether they provide students with sufficient skills to enable them to thrive in the labor market and be able to pay off their student debts (for example, Baum 2011). Simple comparisons of student outcomes between the for-profits and other institutions may be misleading: after all, the for-profits disproportionately attract minority, older, independent, and disadvantaged students. Thus, we assess student outcomes of the for-profits relative to other higher education institutions after adjusting for observable differences in students who have attended different types of schools.

The recent and rapid growth of for-profit colleges means that most of the standard individual-level longitudinal data sets do not identify those who went to

Figure 5
Differences in Three-Year Cohort Default Rate by Type of Institution: 2005 to 2008



	[1]	[2]	[3]	[4]
Demographic controls	No	Yes	Yes	Yes
Financial aid controls	No	Yes	Yes	Yes
Degree types, offerings	No	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
City fixed effects	No	No	Yes	Yes
Limited to open admission	No	No	No	Yes
R^2	0.295	0.428	0.642	0.553
Sample size	14,655	14,655	14,655	9,281

Sources: National Student Loan Data System (NSLDS) and Integrated Postsecondary Education Data System (IPEDS).
Notes: Each bar gives the coefficient on a type of institution from a regression where the dependent variable is the three-year cohort default rate for an institution-year observation and the omitted group is four-year public institutions. The sample covers institution-year observations for the fiscal years 2005 to 2008. Demographic controls are fractions part-time, 25 years and older, female, African American, and Hispanic. Financial aid controls are the number of recipients of Pell grants and subsidized and unsubsidized federal loans, total yearly disbursement amounts for each, and total loans and Pell grants per enrollee. Degree types and offerings are indicators for distance education, remedial course offerings, whether the institution offers assistance with job placement, whether it offers part-time employment services for enrolled students, the highest award or degree offered by the institution, and whether it has open admissions. Standard errors are clustered by institution. Table 1 in an online Appendix, available with this paper at ⟨http://e-jep.org⟩, provides the full regression, standard errors, and the effect of separating the for-profits into the subcategories of independents, regional chains, national chains, and online institutions.

for-profit institutions or do not have large enough samples of for-profit students for a meaningful analysis. To overcome these constraints we use the most recent cohort of the Beginning Postsecondary Students Longitudinal Study, known as BPS:04/09. A sample of 2003–2004 first-time beginning postsecondary students are followed, in their first, third, and sixth years since entering an undergraduate institution, up through 2009. Because it covers a recent cohort, a significant fraction of the sample initially enrolled in a for-profit institution. The BPS has detailed student background variables, low attrition rates, and an oversample of students at for-profit institutions yielding approximately 1,950 students starting at for-profits out of a total of about 16,680 students in our main sample.[8]

The Beginning Postsecondary Students data is representative of first-time postsecondary students (those starting an undergraduate program with no previous postsecondary schooling). But because a large fraction of students in for-profit institutions are older, nontraditional students returning to higher-education, they will not be picked up in this sample. Thus, our analysis estimates the for-profit school treatment effect (relative to other types of institutions) for first-time postsecondary students but not for the large group of returning students.

The outcome variables in the Beginning Postsecondary Students data are divided into two major groups. Those concerning college costs and financial aid are given in Table 2, and those regarding student persistence, educational attainment, employment, earnings, and satisfaction with the program are in Table 3. The raw data, given in columns 1–3 of Tables 2 and 3, reveal that beginning postsecondary students at for-profits accumulate larger student debt burdens, are more likely to default on their student loans, have poorer employment outcomes five years after entering postsecondary school, and are less likely to be satisfied with their course of study than students starting at public or private nonprofit schools. The short-run (one-year) dropout rate is slightly lower for starting for-profit students than those starting in a community college. For-profit students in certificate and AA programs have higher completion rates than community college students. In contrast, BA completion rates of for-profit students are much lower than of those starting in four-year public and nonprofit schools.

Using the Beginning Postsecondary Students data, we assess whether the raw mean student outcome differences have been overstated because for-profit students differ from those in the public and the private nonprofit sectors (as was demonstrated in the bottom panel of Table 1). To do this, we adjust the raw outcomes for differences in baseline observables between for-profit students and others using two methods.

[8] We use the sampling weights from the Beginning Postsecondary Students data in all our analyses to account for the variation in sampling rates among different student subgroups. The attrition rates from the BPS:04/09 by the final 2009 survey round are relatively balanced by starting institution at 6.4 percent for students from for-profits, 10.9 percent for community college students, and 10.7 percent for students from four-year public and nonprofit schools. The differences in attrition rate by starting institution type are small and not statistically significant after conditioning on baseline covariates. Unweighted sample sizes are rounded to the nearest 10.

Table 2
Differences in College Costs and Financial Aid between For-Profit Institutions and Other Schools for First-Time Undergraduates: 2004/2009 Beginning Postsecondary Students Longitudinal Study

Dependent variables	*Beginning postsecondary students (full sample)*				
	Dependent variable means			*For-profit institution impact*	
	4-year public and nonprofits (1)	*2-year public and nonprofits (2)*	*For-profits (3)*	*OLS (4)*	*Matching (5)*
College costs and financial aid, 2003–2004					
Applied for aid (share)	0.895	0.749	0.986	0.094 (0.010)	0.072 (0.011)
Title IV loan and grant aid ($)	3,837	1,022	6,852	4,439 (183)	3,417 (164)
Tuition ($)	9,230	1,269	8,434	5,632 (173)	5,108 (201)
Net tuition minus grants ($)	5,183	734	5,573	4,521 (157)	4,418 (158)
Pell grant ($)	0.285	0.294	0.790	0.190 (0.014)	0.061 (0.020)
Pell grant amount ($)	771	633	2,149	557 (48)	180 (68)
Financial aid through 2009					
Cumulative Pell grant ($)	2,923	2,399	4,084	–170 (146)	–852 (223)
Cumulative Title IV borrowing ($)	8,702	3,502	7,699	3,960 (421)	2,239 (381)
Title IV loan balance in 2009 ($)	8,024	3,306	7,460	4,071 (460)	2,242 (401)
Repaid any amount on loan, conditional on a student loan (share)	0.642	0.640	0.529	–0.093 (0.029)	–0.040 (0.046)
Defaulted on loan, conditional on a student loan (share)	0.035	0.056	0.188	0.067 (0.018)	0.082 (0.018)
Sample size	8,760	5,970	1,950		

Source: BPS:04/09 Restricted-Use Data File. BPS:04/09 is Beginning Postsecondary Students Longitudinal Study data for 2003–2004 first-time beginning postsecondary students in their first, third, and sixth years since entering an undergraduate institution, through 2009.
Notes: The ordinary least squares (OLS) column reports coefficient estimates (robust standard errors) for a for-profit institution dummy variable in regressions for each dependent variable, estimates that include the following covariates: dummy variables for race, sex, citizenship, born in the United States, parents born in the United States, English as the native language, household size, distance of school from home, lives with parents, marital status, single parenthood, independent student, number of kids, use of child care, maternal and paternal education categories, high school diploma, GED receipt, delayed enrollment after high school, certificate or degree program, degree expectations, region, and on- or off-campus residence; and second-order polynomials in age, prior income (own for independent students and family for dependent students), household income percent of the poverty line, expected family contribution from the FAFSA (Free Application for Federal Student Aid), individual adjusted gross income from tax returns and government transfers. Each number in the "Matching "column represents the average treatment on the treated estimate (standard error) for going to a for-profit institution using from nearest neighbor (propensity score) matching with replacement and excluding observations outside of common support. The same covariates used in the ordinary least squares regressions were used for the matching models. The ordinary least squares and matching model estimates use the BPS sampling weights. Unweighted sample sizes are rounded to the nearest 10.

Table 3

Differences in Student Outcomes between For-Profit Institutions and Other Schools for First-Time Undergraduates: 2004/2009 Beginning Postsecondary Students Longitudinal Study

Dependent variables	*Beginning Postsecondary Students (full sample)*				
	Dependent variable means			*For-profit institution impact*	
	4-year public and nonprofits (1)	*2-year public and nonprofits (2)*	*For-profits (3)*	*OLS (4)*	*Matching (5)*
Persistence and educational attainment					
Left school in 2003–2004 (share)	0.062	0.233	0.212	–0.046	–0.051
				(0.016)	(0.018)
Attained certificate (if enrolled in certificate program; share)	—	0.424	0.537	0.086	0.046
				(0.036)	(0.034)
Attained AA (if enrolled in AA program; share)	—	0.224	0.284	0.041	0.019
				(0.028)	(0.029)
Attained AA or more (if enrolled in AA program; share)	—	0.283	0.291	–0.006	–0.016
				(0.028)	(0.030)
Attained BA (if enrolled in BA program; share)	0.658	—	0.262	–0.115	–0.194
				(0.045)	(0.052)
Idle (not employed, not enrolled) at 2009 survey (share)	0.106	0.133	0.236	0.052	0.058
				(0.017)	(0.017)
Enrolled in 2009 (share)	0.271	0.389	0.216	–0.114	–0.080
				(0.018)	(0.019)
Employment and earnings (for those no longer enrolled in 2009)					
Any job in 2009 (share)	0.839	0.784	0.706	–0.028	–0.031
				(0.021)	(0.022)
Earnings from work in 2009 ($)	28,613	24,795	19,950	–1,771	–1,936
				(931)	(950)
Earnings from work in 2009, conditional on employment ($)	34,080	31,622	28,243	–1,355	–243
				(934)	(937)
Unemployed and seeking work (share)	0.121	0.148	0.232	0.048	0.067
				(0.019)	(0.020)
Unemployed 3 months or more after leaving school (share)	0.238	0.259	0.404	0.077	0.084
				(0.022)	(0.023)
Earnings less than gainful employment standard (share)	0.135	0.046	0.271	0.194	0.147
				(0.019)	(0.017)
Course content and job and school satisfaction					
Remedial coursework in 2003–4 (share)	0.181	0.289	0.076	–0.180	–0.187
				(0.015)	(0.017)
Left school because dissatisfied (2003–2004) (share)	0.012	0.024	0.081	0.043	0.048
				(0.009)	(0.009)
Left school because dissatisfied (2003–2006) (share)	0.032	0.051	0.117	0.052	0.053
				(0.013)	(0.011)
Education was worth the cost (share)	0.802	0.821	0.648	–0.204	–0.179
				(0.019)	(0.017)
Loans were a worthwhile investment (share)	0.836	0.803	0.664	–0.143	–0.121
				(0.022)	(0.024)
Satisfied with major or program (share)	0.860	0.871	0.789	–0.097	–0.065
				(0.017)	(0.015)
Satisfied with current job, (employed, not enrolled; share)	0.772	0.764	0.752	–0.011	–0.032
				(0.025)	(0.023)
Sample size	8,760	5,970	1,950		

Source and Notes: See Table 2.

The first method is a standard ordinary least squares regression of student outcomes on a rich set of covariates of student baseline characteristics at entry into college (listed in the table notes), and a dummy variable for starting postsecondary schooling in a for-profit institution. The alternative method is a matching approach, which takes students starting in for-profits as the treatment group and students starting in public and private nonprofit schools as the control group. We compare the outcomes of the for-profit students to the control group members who are observably comparable to for-profit students. More specifically, we estimate the average treatment-on-treated effect of starting in a for-profit institution using nearest neighbor (propensity score) matching models with replacement excluding observations outside of common support.[9] For educational attainment outcomes, the estimation samples are separated into the subgroups of students initially enrolled in each type of program (certificate, AA, BA).

The ordinary least squares results are shown in column 4 for the full sample and those for the matching estimator are in column 5 of Tables 2 and 3. The ordinary least squares and matching approaches produce qualitatively and quantitatively similar estimates for almost every outcome considered.

Our conclusions with regard to the relative performance of students starting in for-profit institutions are mixed. For-profit students have a higher probability of staying with a program through its first year. Early persistence translates into a higher probability of obtaining a degree or certificate in a one- or two-year program. The ordinary least squares estimates indicate that certificate seekers starting at for-profits are almost 9 percentage points more likely to gain a certificate than community college students. Although for-profit students seeking an AA are somewhat more likely than community college students to attain an AA degree, they are less likely to continue to higher-level college courses and to gain a BA degree. The matching estimates indicate that the for-profit advantage in completing certificate and AA programs is more modest and less statistically significant than the ordinary least squares estimates.

Students in for-profit institutions are also much less likely to report taking remedial courses in their first year in postsecondary school than students in other institutions. The greater ability of for-profit students to take courses they consider directly relevant and not languish in remedial courses may play a role in their greater first-year retention rates.[10]

For the longer undergraduate programs, such as BA, for-profits do not fare as well as four-year public and private nonprofit institutions. The ordinary least squares estimate implies a 12 percentage point completion deficit and the matching

[9] We implement the nearest-neighbor matching estimator in STATA using the routines developed by Becker and Ichino (2002).

[10] See Rosenbaum, Deil-Amien, and Person (2006) for rich case study evidence of the roles of clearer program paths, more relevant courses, and student services in better retention and short program completion rates for students in for-profit schools relative to community colleges. Rutschow and Schneider (2011) summarize recent evidence from interventions designed to improve students' progress through remedial courses at community colleges.

model implies a 19 percentage point deficit for students starting BA programs at for-profits. The control group of students in the full range of public and private nonprofit four-year schools is probably less comparable in the case of BA students than for certificate and AA programs. But even when the sample is restricted to students starting in nonselective schools, a statistically significant deficit of almost 5 percentage points remains (details in Appendix Table 2, available online with this paper at ⟨http://e-jep.org⟩).[11]

Also, for-profits leave students with considerably higher debt, even conditional on a rich set of observables. For-profit students face higher sticker-price tuition and pay higher net tuition (tuition plus fees minus grants) than comparable students at other institutions. Students who began at a for-profit school default on their loans at higher rates than other students conditional on controls for demographics, academic preparation, and pre-enrollment family resources. For-profit students have substantially higher default rates even when comparing students across school types with similar cumulative debt burdens. For example, the default rate by 2009 for the BPS:04/09 students with $5,001 to $10,000 in cumulative federal student loans is 26 percent for students from for-profits versus 10 percent for those from community colleges and 7 percent for those from four-year public and nonprofit schools; and for those with $10,001 to $20,000 in debt, the default rate among for-profit students is 16 percent versus a 3 percent rate for community college students and 2 percent rate for other four-year college students.

Although the vast majority of students from for-profits express satisfaction with their course of study and programs, they report significantly lower satisfaction than observably similar students starting in public and nonprofit schools. Students who began in for-profit colleges are also less likely to state that their education was worth the amount they paid and are less apt to think their student loans were a worthwhile investment. Even though the for-profits have higher short-run retention of students, their students are more likely to leave their certificate or degree programs before completion because of dissatisfaction with the program.

In terms of economic outcomes in the medium-run, for-profit students are more likely to be idle (that is, not working and no longer enrolled in school) six years after starting college. Among the students who left school by the 2009 wave of the BPS survey, those from for-profits are more likely to be unemployed and to have experienced substantial unemployment (more than three months) since leaving school. For-profit students no longer enrolled in 2009 have earnings from work in 2009 that are $1,800 to $2,000 lower (or 8 to 9 percent of their predicted mean earnings) than had they gone to another type of institution.[12] Some of the earnings

[11] In addition, Appendix Tables 3 to 5 present comparable analyses for the full range of student outcomes for the subsamples of Beginning Postsecondary Students starting certificate programs, AA programs, and BA programs respectively.

[12] In slight contrast, Cellini and Chaudhary (2011) find similar weekly earnings gains of around 6 percent to attending a two-year AA program at a private or public two-year college and of 15 to 17 percent (or 8 percent per year of education) to completing an AA degree at private postsecondary institutions (largely for-profit schools) and at public institutions (largely community colleges) using an individual fixed effects

reduction is due to lower rates of employment. Once we condition on employment, for-profit students have modestly lower earnings and slightly lower job satisfaction, but neither difference is statistically significant.

For-profit schools, therefore, do better in terms of first-year retention and the completion of shorter certificate and degree programs. But their first-time postsecondary students wind up with higher debt burdens, experience greater unemployment after leaving school and, if anything, have lower earnings six years after starting college than observationally similar students from public and nonprofit institutions. Not surprisingly, for-profits students end up with higher student loan default rates and are less satisfied with their college experiences.

Lower satisfaction with the programs may provide an additional psychological factor accounting for the high default rates of for-profit students, even for those with modest absolute student debt levels. In fact, students in this dataset from for-profits with less than $2,500 in federal student loan debt had a default rate of 20 percent by 2009 as compared with 12 percent for students from community colleges and 4 percent for those from four-year public and nonprofit institutions. These patterns are troubling since the consequences of federal student loan default cannot be escaped through bankruptcy and can adversely impact an individual's credit rating and future access to credit, not to mention result in wage garnishment, harassment by private collection agencies, and tax refund offsets.

Although we have used the detailed background covariates in the Beginning Postsecondary Students survey data to make comparisons between individuals who are as similar as can be observed, we do not have quasi-experimental variation concerning who goes to which type of higher-education institution. Thus, one needs to be cautious in providing a causal interpretation of the estimated for-profit school treatment effects in Tables 2 and 3 since the potential problem of selection bias from nonrandom sorting on unobservables remains. Furthermore, our comparison of the medium-term outcomes for beginning postsecondary students starting at for-profits versus comparable students starting at other higher-education institutions does not directly provide information on whether attendance at a for-profit college (or, for that matter, attendance at public or private, nonprofit colleges) is a worthwhile (private or social) investment.

Nimble Critters or Agile Predators?

The U.S. economy has experienced a substantial increase in the pecuniary returns to postsecondary education since 1980, particularly for BA and higher degrees (Autor, Katz, and Kearney 2008; Goldin and Katz 2008). At the same time,

strategy of comparing earnings before and after college using workers under 30 years old in the 1997 National Longitudinal Survey of Youth. Cellini and Chaudhary likely understate the relative economic returns to going to a public two-year college relative to a private for-profit institution by dropping from their sample the students who continued beyond an AA to get a BA or more.

state budgetary difficulties have constrained the expansion of public sector higher education; for example, Cellini (2009) provides compelling evidence from California on how public sector funding constraints on community colleges increased the rate of entry of for-profit colleges. In the meantime, federal and state financial aid for students going to for-profit institutions has become more available and generous (for example, Cellini 2010). Based on these factors, and others discussed in this paper, the for-profit postsecondary school sector became the fastest growing part of U.S. higher education from the 1990s through 2010. Increased regulatory scrutiny and adverse publicity from Congressional hearings, investigative reporting, and Government Accountability Office (GAO) audits have led to a substantial slowdown in the growth of for-profit enrollments in 2011 and actual declines in new students at many of the larger national chains (Steinerman, Volshteyn, and McGarrett 2011; Fain 2011).

Evaluating the successes and failures of U.S. for-profit higher education must go beyond mean outcomes and consider the distribution of labor market effects and financial default rates. For many, the for-profits have been a success. They have played a critical role in expanding the supply of skilled workers in an era of tight state budgets and stagnating state appropriations to public sector schools. They have provided educational services to underserved populations. Their innovative use of web services has further allowed them to accommodate nontraditional students. Their disproportionate share of federal student grants and loans has enabled them to provide skills to disadvantaged populations. Short-run retention is high and the for-profits do an admirable job of graduating students from shorter certificate programs. The vast majority of their students are satisfied with their programs.

But the for-profits also charge higher tuition and fees than public sector alternatives, and their students are more likely to end up unemployed and with substantial debts. Students who attended a for-profit have much higher default and nonrepayment rates on federal student loans than do observationally similar students who attended a public or private nonprofit institution.

The U.S. Department of Education (2011b) has recently sought to address this issue of the high default rate on loans to students at for-profit institutions by passing "Gainful Employment" regulations, which will require most for-profit programs and certificate programs at public and nonprofit institutions to pass at least one of three metrics to remain Title IV–eligible: 1) at least 35 percent of former students repaying their loans ("repaying" defined as reducing their loan by at least $1 over the course of a year); 2) annual loan payments not exceeding 30 percent of a typical graduate's discretionary income; or 3) annual loan payments not exceeding 12 percent of a typical graduate's earnings.

How these rules will work in practice, as students and for-profit institutions adjust to their presence, remains to be seen. The former students of for-profit institutions have comparable (but slightly lower) earnings, combined with substantially higher loan burdens, relative to other school leavers, suggesting that some for-profit institutions may face challenges meeting the new Gainful Employment standards. As one example, consider the rule that the debt burden (annual federal student loan yearly payments) should not exceed 12 percent of annual earnings for a typical graduate. In fact we find (conditional on observables), in Table 3 for the Beginning Postsecondary Students data, that for-profit students would have had a 15 to 19 percentage point lower rate of meeting the recently enacted Gainful Employment earnings threshold in 2008 (four to five years after starting) than would students from other types of institutions.

In effect, the Gainful Employment rule seeks to hold the for-profits more accountable and put a greater burden on the schools, rather than only on the students who have difficulties in repaying their loans. The new regulations will also require institutions to disclose their program costs, as well as completion, placement, and loan repayment rates. These regulations will increase transparency but may be insufficient to contain an agile predator. A reality check by a third party might be needed before a student is allowed to take out a loan.

The for-profits have taken a large burden of increased enrollment in higher education off the public sector. The high default rates of their students on federal loans, however, increase their cost to the taxpayer. Regulating for-profit colleges is tricky business. The challenge is to rein in the agile predators while not stifling the innovation of these nimble critters.

■ *We gratefully acknowledge the superb research assistance of Tanya Avilova, Jason Poulos, and Bernie Ziprich. We are grateful to Stephanie Cellini and the Editors for helpful comments. The research reported here was supported by the Institute of Education Sciences, U.S. Department of Education, through Grant R305C110011 to Teachers College, Columbia University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.*

References

Autor, David H., Lawrence F. Katz, and Melissa S. Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists." *Review of Economics and Statistics* 90(2): 300–323.

Baum, Sandy. 2011. "Drowning in Debt: Financial Outcomes of Students in For-Profit Colleges." Testimony to the U.S. Senate Committee on Health, Education, Labor and Pensions, June 7.

Becker, Sascha O., and Andrea Ichino. 2002. "Estimation of Average Treatment Effects based on Propensity Scores." *STATA Journal* 2(4): 358–77.

Bennett, Daniel L., Adam R. Lucchesi, and Richard K. Vedder. 2010. "For-Profit Higher Education: Growth, Innovation, and Regulation." A Policy Paper from the Center for College Affordability and Productivity, July. http://heartland.org/sites/all/modules/custom/heartland_migration/files/pdfs/29010.pdf.

Breneman, David W., Brian Pusser, and Sarah E. Turner, eds. 2006. *Earnings from Learning: The Rise of For-Profit Universities.* Albany, NY: State University of New York Press.

Cellini, Stephanie Riegg. 2009. "Crowded Colleges and College Crowd-Out: The Impact of

Public Subsidies on the Two-Year College Market." *American Economic Journal: Economic Policy* 1(2): 1–30.

Cellini, Stephanie Riegg. 2010. "Financial Aid and For-Profit Colleges: Does Aid Encourage Entry?" *Journal of Policy Analysis and Management* 29(3): 526–52.

Cellini, Stephanie Riegg, and Latika Chaudhary. 2011. "The Labor Market Returns to a Private Two-Year College Education." http://home.gwu.edu/~scellini/Index/Research_files/Cellini%26Chaudhary_Returns_April11.pdf.

Cellini, Stephanie Riegg, and Claudia Goldin. Forthcoming. "Does Federal Student Aid Raise Tuition?: New Evidence on For-Profit Colleges" NBER Working Paper.

Chronicle of Higher Education. 2010. "CEO Compensation at Publicly Traded Higher-Education Companies." June. 23. http://chronicle.com/article/Graphic-CEO-Compensation-at/66017/.

Fain, Paul. 2011. "Enrollments Tumble at For-Profit Colleges." *Inside Higher Ed,* November 11. http://www.insidehighered.com/news/2011/11/11/enrollments-tumble-profit-colleges.

Gibson, Ellen. 2009. "College Campuses Debate Administrators' Lofty Pay." *Bloomberg Business Week*, February 16. http://www.businessweek.com/bwdaily/dnflash/content/feb2009/db20090216_614557.htm?chan=top+news_top+news+index+-+temp_news+%2B+analysis.

Goldin, Claudia, and Lawrence F. Katz. 2008. *The Race between Education and Technology*. Cambridge, MA: Belknap Press, Harvard University Press.

Health, Education, Labor and Pensions Committee, U.S. Senate. 2010. "Benefiting Whom? For-Profit Education Companies and the Growth of Military Educational Benefits." December 8. http://harkin.senate.gov/documents/pdf/4eb02b5a4610f.pdf.

Hentschke, Guilbert C. 2010. "Innovations in Business Models and Organizational Cultures: The For-Profit Sector." Unpublished paper, USC Rossier School of Education, June.

Knapp, Laura G., Janice Kelley-Reid, and Scott A. Ginder. 2011. *Postsecondary Institutions and Price of Attendance in the United States: 2010–11: Degrees and Other Awards Conferred: 2009–10, and 12-Month Enrollment: 2009–10. First Look*. NCES 2011–250. Washington, D.C: U.S. Department of Education. http://nces.ed.gov/pubs2011/2011250.pdf.

Rosenbaum, James E., Regina Deil-Amien, and Ann E. Person. 2006. *After Admission: From College Access to College Success*. New York: Russell Sage Foundation Press.

Rutschow, Elizabeth Zachry, and Emily Schneider. 2011. *Unlocking the Gate: What We Know About Improving Developmental Education*. New York: Manpower Demonstration Research Corporation.

Steinerman, Andrew, Jeffrey Volshteyn, and Molly McGarrett. 2011. *Education Services Data Book*. September. J.P. Morgan, North American Equity Research, Business and Education Services.

Turner, Sarah E. 2006. "For-Profit Colleges in the Context of the Market for Higher Education." In *Earnings from Learning: The Rise of For-Profit Universities,* edited by D. Breneman, B. Pusser, and S. Turner, 51–68. Albany: State University of New York Press.

U.S. Department of Education. 2011a. *2011–2 Federal Student Aid Handbook, Vol. 2, School Eligibility and Operations*. http://ifap.ed.gov/fsahandbook/1112FSAHbkVo2.html.

U.S. Department of Education. 2011b. "Additional Background on the Gainful Employment Regulations." Go to: http://www.ed.gov/news/press-releases/gainful-employment-regulations, click on "Supplemental Information."

U.S. Department of Education, Federal Student Aid Data Center. 2011. *Proprietary School 90/10 Revenue Percentages,* Report for 2009–2010 Award Year. Available at http://federalstudentaid.ed.gov/datacenter/proprietary.html.

U.S. Department of Education, National Center for Education Statistics (NCES). 2010. *Digest of Education Statistics: 2010*. http://nces.ed.gov/programs/digest/.

U.S. Department of Education. National Center for Education Statistics (NCES). 2011. *The Condition of Education 2011*. http://nces.ed.gov/programs/coe/.

U.S. Department of Education, National Center for Education Statistics (NCES), Integrated Postsecondary Education Data System (IPEDS). 2009. IPEDS Fall 2009 Compendium Tables. http://nces.ed.gov/das/library/tables_listings/fall2009.asp.

U.S. Government Accountability Office. 2010. "For-Profit Colleges: Undercover Testing Finds Colleges Encouraged Fraud and Engaged in Deceptive and Questionable Marketing Practices." GAO-10-948T, August 4. http://www.gao.gov/products/GAO-10-948T.

Watkins, Barbara L. 1991. "A Quite Radical Idea: The Invention and Elaboration of Collegiate Correspondence Study." In *The Foundations of American Distance Education: A Century of Collegiate Correspondence Study,* edited by Barbara L. Watkins and Stephen J. Wright, 1–35. Dubuque, IA: Kendall/Hunt.

Winston, Gordon C. 1999. "For-Profit Higher Education: Godzilla or Chicken Little?" *Change* 31(1): 12–19.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 165–192

Student Loans: Do College Students Borrow Too Much—Or Not Enough?†

Christopher Avery and Sarah Turner

Total student loan debt rose to over $800 billion in June 2010, overtaking total credit card debt outstanding for the first time. By the time this article sees print, the continually updated Student Loan Debt Clock (at ⟨http://www.finaid.org/loans/studentloandebtclock.phtml⟩) will show an accumulated total of roughly $1 trillion. New federal student loans for higher education amounted to $97 billion in 2009–2010: $66.8 billion to undergraduates and $31 billion to graduate students. Borrowing to finance educational expenditures has been increasing—more than quadrupling in real dollars since the early 1990s—as shown in Figure 1. The sheer magnitude of these figures has led to increased public commentary on the level of student borrowing.

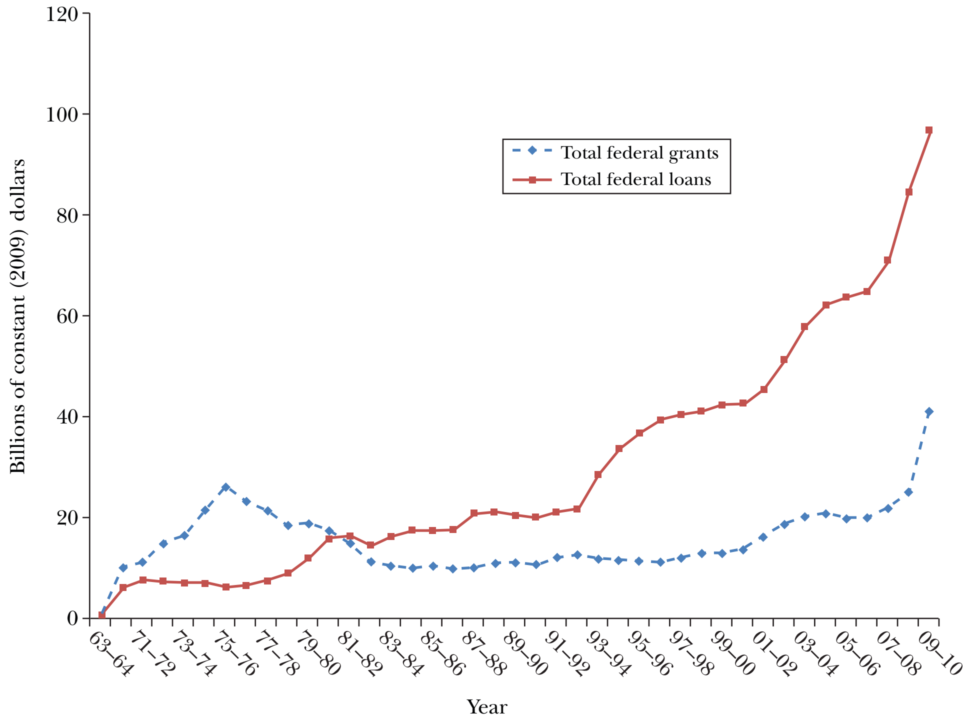
On the one side, it has become fashionable to suggest that we are in the midst of an "education bubble" (for example, Schumpeter Blog 2011). As Surowiecki (2011) summarizes, "[Y]ou can't flip a college degree the way you can flip a stock, or even a home. But what bubble believers are really saying is that young people today are radically overestimating the economic value of going to college, and that many of them would be better off doing something else with their time and money." Similarly, Kamenetz (2006) argues that a combination of wage declines in entry-level jobs and increases in college tuition have placed many high school graduates in a no-win position, pressuring them to take on unmanageable levels of financial risk in the form of student loans. Of course, the depressed job market during the

■ *Christopher Avery is Roy E. Larsen Professor of Public Policy and Management, Kennedy School of Public Policy, Harvard University, Cambridge, Massachusetts. Sarah Turner is University Professor of Economics and Education, University of Virginia, Charlottesville, Virginia. Their e-mail addresses are ⟨chris_avery@harvard.edu⟩ and ⟨sturner@virginia.edu⟩.*

† To access the Appendices, visit http://dx.doi.org/10.1257/jep.26.1.165.

doi=10.1257/jep.26.1.165

Figure 1
Trends in Federal Grant and Loan Aid



Source: College Board, *Trends in Student Aid*, 2010.

Great Recession and its aftermath has only strengthened such concerns, and added others. Rothstein and Rouse (2011) provide evidence that high debt burdens make students less likely to choose a lower-paying career, like becoming a teacher. Gicheva (2011) suggests that additional student debt of $10,000 decreases the long-term probability of marriage by 7 percentage points. A 2010 poll found that 85 percent of college graduates were planning to move back home after graduation (Dickler 2010). Newspaper stories tell of students who finish their undergraduate degree with $100,000 or more in debt (Leiber 2010).

On the other side, the earnings premium for a college degree relative to a high school degree nearly doubled in the last three decades (Goldin and Katz 2008). Further, there is no particular evidence this earnings premium has declined as a result of the Great Recession, as the alternative to a weak labor market for college graduates today is a much weaker labor market for those without a college degree. In November 2011, data from the Bureau of Labor Statistics website shows that the unemployment rate for college graduates (including those with advanced degrees) was 4.4 percent, while high school graduates faced an unemployment rate of 8.5 percent and those with collegiate attainment less than a BA faced an unemployment rate of 7.6 percent. While the fraction of college-educated workers in the labor force has increased considerably in recent decades, current projections suggest the education level of the labor force will increase little, if at all, in the early twenty-first century (Ellwood 2001; see also Goldin and Katz 2008).

Concurrent with the recent and dramatic increase in the college earnings premium, overall undergraduate enrollment in college has increased from 10.5 million in 1980 to 17.6 million in 2009, while the annual volume of federal loans has increased more rapidly from 2.3 million loans in 1980 to 10.9 million loans in 2009 (Institute of Education Sciences 2010; see data at ⟨http://www2.ed.gov/finaid/prof/resources/data/opeloanvol.html⟩). In theory, federal student loans can help to overcome a problem of social underinvestment in capital markets that was described by Milton Friedman in his 1962 *Capitalism and Freedom*:

> This underinvestment in human capital presumably reflects an imperfection in the capital market: investment in human beings cannot be financed on the same terms or with the same ease as investment in physical capital. It is easy to see why there would be such a difference. If a fixed money loan is made to finance investment in physical capital, the lender can get some security for his loan in the form of a mortgage or residual claim to the physical asset itself, and he can count on realizing at least part of his investment in case of necessity by selling the physical asset. If he makes a comparable loan to increase the earning power of a human being, he clearly cannot get any comparable security; in a non-slave state, the individual embodying the investment cannot be bought and sold. But even if he could, the security would not be comparable. The productivity of the physical capital does not—or at least generally does not—depend on the co-operativeness of the original borrower.

In this perspective, student loans can potentially improve the efficiency of the economy by raising the supply of college-educated workers in the labor market. Moreover, because credit constraints are most likely to affect students from low-income families, student loans can reduce both educational and income inequality among those in the same generation and between generations. Higher levels of federal student loans may also reduce supply constraints generated by declining state-level support for public colleges and universities, reducing the extent to which collegiate attainment is deterred by insufficient educational offerings.

So are college students borrowing too much or too little?[1] The question turns on the source of the college wage premium and on the magnitude of that wage premium for the marginal college student. The college experience provides graduates with skills and social networks, and a college degree may serve as a signal of ability to employers. These factors suggest a causal link between collegiate attainment

[1] Our focus in this analysis is on student borrowing among undergraduates in the United States. We note that loan funding is also an important source of funding in graduate programs, particularly in professional fields, though this is not the focus of our analysis.

and future wages. But it is also possible that students who choose to go to college would still be unusually successful if they entered the workforce directly upon high school graduation—that is, the college wage premium could result primarily from self-selection.

At the individual level, the choice of how much to borrow requires substantial information about expected collegiate attainment and the future path of earnings under alternative educational attainment scenarios. The connection between educational attainment and career success has been sufficiently well-publicized that even disadvantaged students from urban public schools tend to produce relatively accurate estimates of average wages at age 25 for those with and without a college degree (Avery and Kane 2004; Rouse 2004; see also Dominitz and Manski 1996). Yet, it is far from clear that young people are able to estimate *their own* future earnings accurately and take into account the extent to which there may be systematic differences between expectations and realizations. Manski (1993) emphasizes the importance of understanding how youth form expectations about future earnings and whether they condition on ability, in predicting their educational attainment and their own returns to education. Moreover, on the cost side, researchers have documented that students often misunderstand financial aid packages, fail to understand the much greater cost of consumer loans (such as credit card debt) relative to student loans, and miscalculate the trade-off between academic study and market work (Long 2004; Burdman 2005; Somers, Woodhouse, and Cofer 2004; King 2002; St. John 2004; Warwick and Mansfield 2000). Information constraints may lead to underborrowing if students do not avail themselves of borrowing opportunities, or to overborrowing if students overestimate the return to education.

Our focus in this paper is to move the discussion of student loans away from anecdote and to establish a framework for considering the use of student loans in the optimal financing of collegiate investments. We begin by providing a brief summary of the institutional framework and broad trends associated with U.S. student lending. The next section turns to the consideration of an analytic framework for determining how much a student should be willing to borrow and how this sum has likely changed over time. We will emphasize considerations of uncertainty and heterogeneity: even if the return to college is favorable on average, it need not be favorable for all agents. In our terminology, uncertainty is unknown, while heterogeneity represents difference among agents in their personal returns to college—including nonmonetary returns—that are known (or knowable) to them. We then look to available—albeit limited—evidence to assess which types of students are likely to be borrowing too much or too little.

Borrowing for College

Federal Student Lending Programs

There are currently four major federal sources of loans for higher education: subsidized Stafford loans, the unsubsidized Stafford loans, the Parent Loans for

Undergraduates (PLUS) program, and the Perkins loans program. We provide a brief overview of these programs, along with a comparison to private sector loans for higher education.

The Higher Education Act of 1965 created the Stafford loan program, which has long been by far the largest federal student loan program. Federal student loans were initially means tested and have traditionally featured favorable terms for students from poor families through the subsidized Stafford Loan program. These loans offer three substantive advantages over private market loans: 1) subsidized interest rates; 2) deferral of repayment while student is enrolled at least half-time in college; and 3) subsidies for interest payments while a student is enrolled at least half-time in college. These subsidized Stafford loans rose from about $15 billion in 1990 to $20 billion in 2000, before jumping to $35 billion in 2009 (all in constant 2009 dollars).

In 1992, Congress created an unsubsidized Stafford program for borrowers ineligible for the means-tested subsidized Stafford loans. In 2011–12, for example, subsidized Stafford loans carry an interest rate of 3.4 percent, but the unsubsidized Stafford loans carry an interest rate of 6.8 percent. Annual new loans in the unsubsidized Stafford program had already reached $15 billion by 2000, but since then have leaped to almost $45 billion in 2009. In addition, the federal government introduced a student loan program for parents in 1980 called Parent Loans for Undergraduate Students Program (PLUS). This program loaned about $2 billion in 1990 and $5 billion in 2000, before rising to $12 billion in 2009.

Finally, the 1958 National Defense Education Act created the National Defense Student Loan Program (NDSL) which is now known as the Federal Perkins Loan Program. Perkins loan funds have been distributed by the federal government to collegiate institutions, with institutions in turn allocating funds on the basis of financial need. In the 2009–10 academic year, about 520,000 students from 1,800 institutions received Perkins loans, averaging $2,125, so total spending on the program is a little over $1 billion. The Perkins program is set to expire in 2012, limiting new loans to any funds available from repayments.

Overall, in 2009, subsidized Stafford loans accounted for about 43 percent of federal loan volume, with unsubsidized Stafford loans accounting for 40 percent and PLUS loans for 16 percent. However, it is naturally important to remember that federal lending for higher education is not a comprehensive measure of total lending for that purpose. For example, the growth in federal student loans may overstate the true increase in borrowing for students to attend college if the increase in Stafford loans supplanted other types of loans—like home equity loans in some cases—used previously to pay for college costs.

Starting in the mid-1990s, there has also been a dramatic increase in private sector loans that can be explicitly linked to higher education, driven in part by increased demand for such loans and in part by financial services sector innovations such as greater securitization of student loans through asset-backed securities. While private sector loans were about $1.5 billion (constant 2009 dollars) in 1995–96, they grew to $21.8 billion by 2007–2008, representing about 20 percent of all loan funds distributed (College Board 2010a). Because these loans generally carry somewhat higher interest rates than federal loans, students typically take these loans after exhausting other sources of credit. Mazzeo (2007) reviewed private student loan offerings and noted that many of these loans are marketed as supplements to Stafford loans. Mazzeo also suggests that some parents may prefer private lending options over PLUS loans, because the private loans are made in the student's name. Because private lenders have a greater capacity to discriminate among borrowers by their choice of collegiate investments, higher-ability students and students enrolled in the most remunerative degree programs will be offered more credit by private lenders (Lochner and Monge-Naranjo 2011). In the wake of the financial crisis of 2008 and 2009—and its effects on the market for securitized loans in general—private sector student loans have returned to their historical level of about 7 percent of the market.

Shifting from borrowing to repayment, conventional student loans carry monthly payments over a 10-year horizon. With federal loans, students can choose from among alternative repayment options, which may increase the duration of the loan to 25 years, and graduated payments, with payments increasing every two years (see Krueger and Bowen, 1993, for discussion of income-contingent repayment plans).

Student Borrowing

To be eligible for federal loan options, a student must complete the Free Application for Federal Student Aid (FAFSA) form (available at ⟨http://federalstudentaid.ed.gov⟩).This application qualifies students for federal student aid programs authorized under Title IV of the Higher Education Act, including both direct loan programs and Pell grants. Eligibility for student loans is restricted to U.S. citizens, permanent residents, and eligible noncitizens (like those granted asylum) with high school degrees or who have passed the General Educational Development (GED) test. Eligibility for subsidized government loans is further restricted to students with demonstrated unmet financial need—or those students for whom cost of attendance minus grant aid minus the "Expected Family Contribution" (calculated through analysis of income and assets) is positive. This level of unmet need serves as a cap on the amount that a student will be permitted to borrow through federal loan programs and is the total cost of attendance less any grant aid (federal, state, or institutional). Economic models predict that students will exhaust borrowing from the lowest cost of capital first (subsidized loans, if the student is eligible), followed by unsubsidized government loans and private loans, though such a pattern does not always hold in the data.

Table 1 provides summary statistics for undergraduate borrowing from federal programs over the past 20 years. During this time, the total volume of federal loans has expanded several-fold, but average loan levels per student borrower were largely constant in real terms. That is, the increase in loans disbursed by the federal government is largely due to an expansion in the number of borrowers over time. In addition to an increase in the number of students enrolling in college over time, the proportion

Table 1
Percentage of All Undergraduate Borrowing, by Student and Institution Characteristics

	1989–90	*1992–93*	*1995–96*	*1999–2000*	*2003–04*	*2007–08*
Percent of undergraduates borrowing						
Total	**19%**	**19%**	**25%**	**27%**	**32%**	**35%**
Type of institution						
Public 4-year	19%	23%	37%	39%	43%	41%
Private nonprofit 4-year	31%	34%	47%	49%	53%	54%
Public 2-year	4%	6%	4%	5%	8%	10%
For-profit	63%	47%	59%	74%	76%	88%
Dependency status						
Dependent	18%	20%	31%	34%	36%	36%
Independent	19%	17%	19%	21%	28%	33%

Source: The information in the table is taken from Table 1.1 of *Trends in Undergraduate Stafford Loan Borrowing: 1989–90 to 2007–08* (Wei 2010). The table is based on data from U.S. Department of Education, National Center for Education Statistics, and 1989–90, 1992–93, 1995–96, 1999–2000, 2003–04, and 2007–08 National Postsecondary Student Aid Studies (NPSAS:90, NPSAS:93, NPSAS:96, NPSAS:2000, NPSAS:04, NPSAS:08).
Note: This table includes both subsidized and unsubsidized borrowing from the Stafford program.

of undergraduates who take out student loans has increased, rising from about 19 percent in 1989–90 to about 35 percent in 2007–2008. As shown in Table 1, this increase in borrowing has been somewhat larger among dependent undergraduate students than independent students.[2] Students who begin at two-year public institutions are the least likely to borrow (about 10 percent in 2007–2008) and borrow the lowest average amounts (conditional on borrowing at all). Students at for-profit institutions are the most likely to borrow (88 percent).

This variation in borrowing by type of institution is a function of both the revenue structure of colleges and universities and the extent to which institutions draw students with a high degree of financial need. For-profit institutions depend largely on student tuition and fees and receive about three-quarters of their funding through federal Title IV loans and grants (as Deming, Goldin, and Katz discuss in their paper in this symposium).

There is a structural reason that average federal loan levels per student have been fairly constant in real terms over time: borrowing under the federal loan programs is limited by both cost of attendance (less grant aid) and nominal loan limits associated with the Stafford, Perkins, and PLUS program. For the Stafford program, annual loan limits are defined in terms of year of study and independent status, rising from $3,500 per year for first-year undergraduates to $8,500 for graduate students. The loan limits associated with the Stafford program bind in many cases, with borrowers

[2] For purposes of financial aid awards, an independent student is a student who meets any one of the following: at least 24 years old; married; a graduate or professional student; a veteran; an orphan; a ward of the court; or someone with legal dependents other than a spouse.

often clustered at the maximum loan level. For two decades from 1987 to 2007, loan limits remained fixed in nominal terms: for example, first-year students were limited to borrowing $2,625 from the subsidized Stafford program. In addition to annual limits, there are lifetime limits on subsidized Stafford ($23,000) and unsubsidized Stafford ($31,000 for dependents and $57,500 for independent). With rapid increases in college costs during the 1990s and unchanging loan limits, the share of undergraduate borrowers reaching loan limits increased from 1989–90 until 2007 when loan limits were increased. In 1989–90, 42 percent of subsidized Stafford borrowers were at the maximum, while 17.8 percent of all Stafford borrowers were at the maximum; by 2003–2004, these numbers had risen to 50.3 and 50.6 percent for subsidized Stafford borrowers and all student borrowers, respectively. With Stafford loan limits raised in 2007, the percentage at the maximum for subsidized and unsubsidized Stafford loans fell to 42.4 and 44.1 percent, respectively, in 2007–2008 (Wei 2010).

College as an Investment

The decision as to whether to invest in one's human capital in the form of education requires that an individual compare the present discounted value of benefits—among which are the gains in future earnings as a result of education—to the present discounted value of costs, including tuition, fees, and foregone wages.[3] In this section, we consider the question of the extent to which the monetary returns from college have exceeded the costs over recent decades for an average student. (For discussion of the nonmonetary returns to college—for example, conditioning on wage levels, the extent to which higher educational attainment predicts higher job satisfaction, see the article by Oreopoulos and Salvanes in the Winter 2011 issue of this journal.) In the next section, we focus on uncertainty and heterogeneity across students.

Suppose that two students graduate from high school simultaneously in June 2009, and that one completes college in four years and subsequently earns wages equal to the average for college graduates at each age, while the other enters the job market immediately and earns wages equal to the average for high school graduates at each age. Based on data from the 2009 Current Population Survey, the gap in average earnings between college graduates and high school graduates starts at $7,000 at age 22 ($28,200 for college graduates versus $21,000 for high school graduates), grows steadily from age 22 to 42 and then levels off at later ages. Though at the point of college graduation the fictitious college graduate would be more than $100,000 behind the high school graduate in the present discounted value of net income, the college graduate overtakes the high school graduate at age 34.

[3] With well-functioning capital markets and full opportunities to borrow, the human capital investment decision of how much education to acquire is separable from the consumption and savings choice at each moment in time conditional on expected lifetime earnings (for a formal demonstration, see Lochner and Monge-Naranjo 2011).

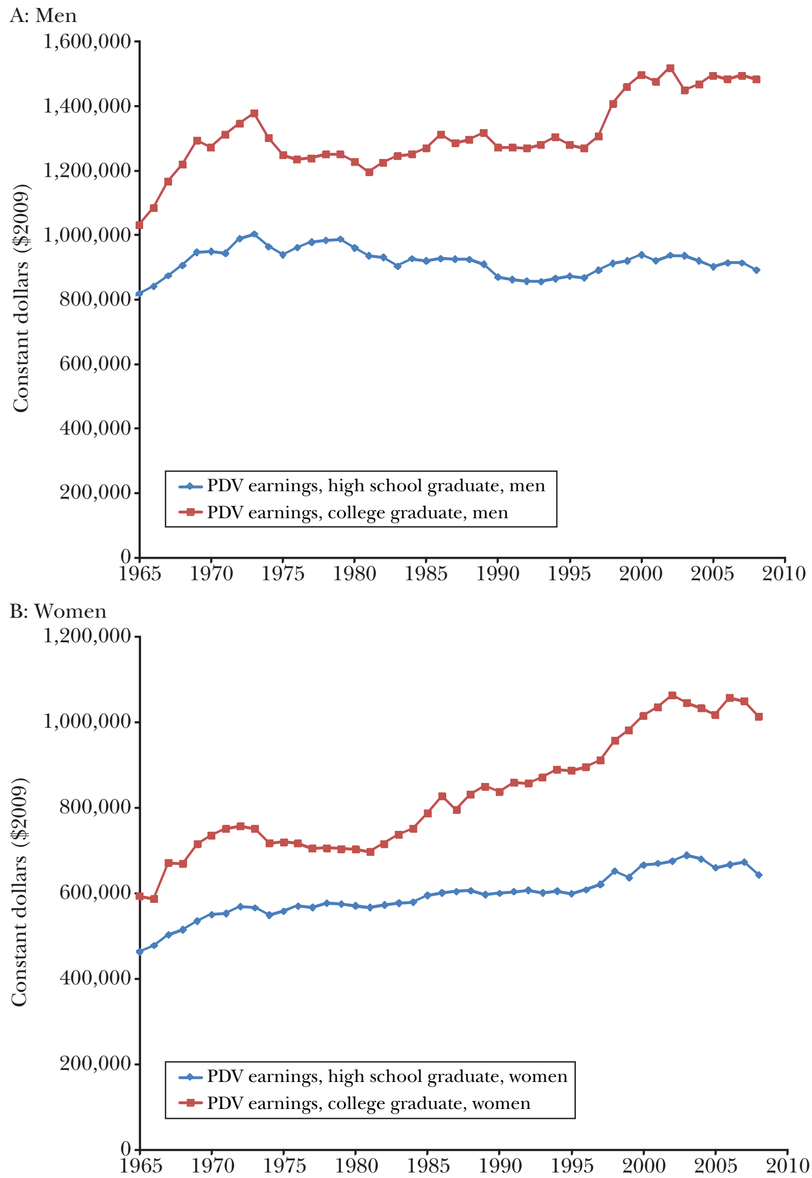
With this example, we want to emphasize that we are not making a causal statement about the magnitude of the returns to education. Such a comparison would rest critically on the assumption that the counterfactual wage distribution for someone earning the average college wage is the average wage of high school graduates. However, at least some of the characteristics that lead a person to select college may also be relevant to their income-earning abilities after college, and observed wages will reflect selection into different levels of education. In addition, changes in earnings streams over time may reflect compositional shifts in the characteristics of individuals with different levels of educational attainment. Later in the paper, we will delve further into these issues of heterogeneity and their implications for the choices of individuals.

In a present discounted value calculation with a 3 percent yearly discount rate, by age 64 the college graduate would have compiled a total of approximately $1.2 million in earnings net of tuition at age 64 as opposed to approximately $780,000 in total earnings for the high school graduate. Of course, this calculation of the average lifetime benefit to a college degree requires a number of assumptions: the discount rate, years of work, growth rate of earnings over the life course, labor force participation, and so forth. But, given the large difference in outcomes between the two fictitious students in this example, the qualitative comparison between them is clearly robust to plausible changes in underlying assumptions. In particular, the comparison is robust to adjustments for the effect of self-selection. For instance, if we assume that half of the difference in wages between a college graduate and a high school graduate is due to self-selection,[4] then the lifetime earnings for the college graduate decline to $925,000 and the college graduate would not overtake the high school graduate until age 42. The present discounted value only becomes the same for the high school graduate and the college graduate if we attribute about 75 percent of the difference in observed earnings to self-selection. This seems like an implausibly large effect given the connection between college graduation and many lucrative career paths.

Figure 2 compares the average lifetime earnings for a college graduate relative to a high school graduate for men and women from 1965 to 2008. The annual values reflect what a man or woman would expect to earn working full time, full year over a career of 42 years, with a discount rate of 3 percent, assuming the college graduate delayed the start of earnings for four years while in school. We calculate the expectation formed in any given year by assuming that the future high school and college graduates will have the future earnings at each age equal to the average earnings of high school and college graduates (respectively) presently observed at each age: thus, the expectation in, say, 1980 is formed based on data across ages for 1980, and so on for each year. The present discounted value of earnings for high school graduates has remained mostly flat (particularly for men). At the same time, the present discounted value of the earnings for a college graduate have risen

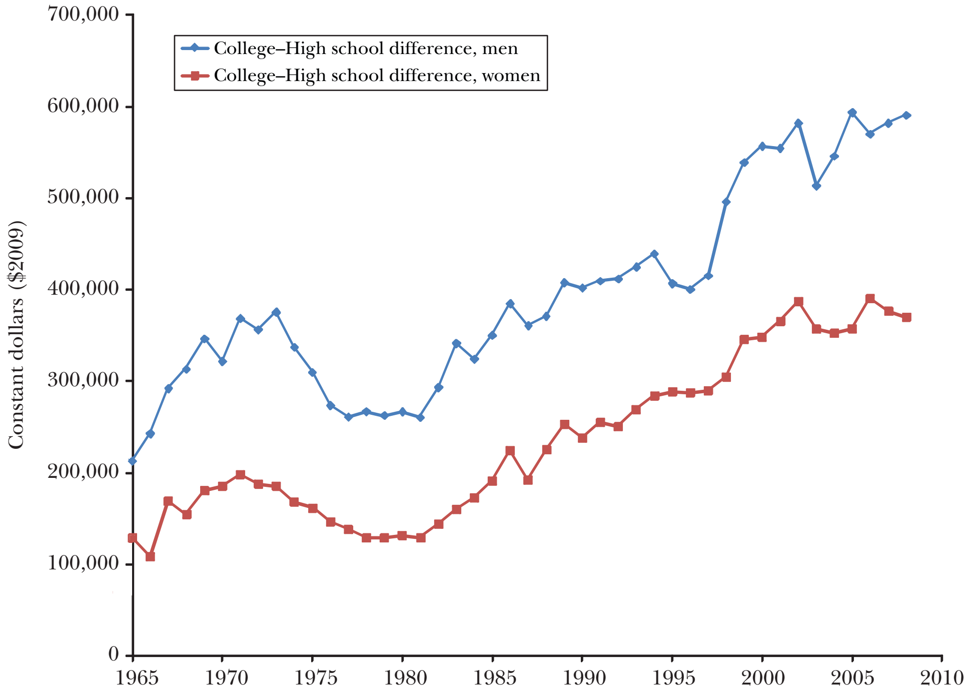
[4] The average yearly earnings at age 22 are $21,000 for a high school graduate and $28,000 for a college graduate; attributing half of this difference to self-selection corresponds to a predicted wage of $24,500 for someone who is switched from being a high school graduate to a college graduate.

Figure 2
Trends in the Present Discounted Value (PDV) of High School and College Earnings Net of Tuition



Notes: Expected earnings are calculated from the March Current Population Survey files for full-time, full-year workers using sample weights, assuming 42 years of work experience per person. Results for college-educated workers are net of four years of tuition and fees associated with appropriate year-specific values for public universities.

Figure 3
Present Discounted Value of College Degree Net of Tuition, 1965–2010



Source: These calculations are based on data from the March Current Population Survey files for full-time, full-year workers using sample weights.
Notes: College–High school difference represents the difference between the present discounted value of the average expected earnings of a college graduate (assuming that earnings begin four years after college entrance and the student pays tuition for four years) and the stream of earnings for a high school graduate. See Figure 2 for earnings calculations.

markedly between 1981 and 2008, rising from $1.2 to $1.5 million for men and from $720,000 to $1.1 million for women.[5]

Figure 3 makes clear that the lifetime earnings increment, on average, of a college degree receipt relative to a high school degree has grown markedly over the last three decades for men and women. These earnings increments are shown in comparison to the discounted value of tuition expenditures over four years (the on-time degree completion for a full-time student) over time in recent years. Thus, even as the present discounted value of tuition for four years at a private college (which would be the most expensive option) has increased over the interval from

[5] These estimates are similar in spirit to Census Bureau estimates produced in Day and Newburger (2002); our estimates of the total value of lifetime earnings to different educational credentials is somewhat lower owing to discounting annual earnings and subtracting expected direct costs of educational investments.

about \$50,000 to \$122,000 (all in constant 2008 dollars), average benefits of college completion in terms of future earnings have increased more rapidly. To be sure, the average net price of college is somewhat below these figures, because grant-based financial aid from government and institutions reduces the price paid by students below the sticker price.[6]

One natural conjecture is that a risk of recession should affect how students invest in a college education, but the direction of this effect is not clear. On one side, the opportunity cost of attending college in terms of foregone wages is lower during a recession, which should tend to increase college attendance during recessions; on the other side, there is a negative effect on wages for those who graduate from college during a recession that can persist as long as ten years (Kahn 2010), which might tend to discourage college attendance in a recession. Figure 3 indicates, however, that the estimated present discounted value of a four-year college degree has increased fairly steadily over the past 30 years through both booms and busts. Further, the comparisons in the figure are based on the average difference in wages for full-time workers with and without BA degrees, but the unemployment rate for college graduates tends to be substantially lower than that for high school graduates in a recession. For that reason, Figure 3 may understate the financial return to a college degree during a recession.[7]

The message is clear: expected lifetime earnings associated with a college degree have increased markedly over time. As the investment value of a college degree rises, it is natural to think of individuals increasing their willingness to borrow to achieve these higher returns.

Of course, a number of factors also affect realized student borrowing which may well diverge from willingness to borrow. For example, the direct cost of college represented by tuition charges has increased markedly in both the public and private sectors, which will tend to increase demand for borrowing among those students who do not receive commensurate increases in financial aid. In addition,

[6] Data from the College Board (2010b, table 7) indicate that tuition and fees net of grant aid changed much less markedly than posted tuition or "sticker price." At private four-year institutions, average net tuition and fees (in 2010 constant dollars) decreased from \$12,230 to \$11,320 between 2000–01 and 2010–11, decreased at public four-year institutions from \$1,990 to \$1,540, and also decreased at public two-year institutions from \$920 to \$670.

[7] At the macroeconomic level, some combination of demographic changes and sectoral shifts in employment would be more likely than a long-lived recession to reduce the financial gains from a college degree indicated in Figure 3. In fact, 35 years ago, in *The Overeducated American*, Richard Freeman noted the dramatic decline in the earnings of new college graduates and argued that there would be little net benefit to further increases in the supply of college graduates. Consistent with Freeman's analysis, Figure 3 suggests that expected financial returns to a college degree were near a long-term low for this time period towards the end of the 1970s. But our computations still indicate a clear positive value for completing college at that time. We compute average lifetime earnings in a given year by simply adding the average earnings of workers at each age in that year. As Smith and Welch (1978) note, although young (age 25 to 34) college graduates were earning relatively low wages, there remained large gaps in wages between college graduates and high school graduates at older ages throughout the 1970s. In essence, the qualitative comparisons from Figure 3 for the 1970s rely on the conjecture that college graduates would continue to enjoy substantial wage gains at age 35 and beyond—a conjecture that has been borne out in subsequent years.

a decline in family resources generated by adverse shocks to parental income or assets could contribute to increased student borrowing. On the other side, a student might react to greater availability of student loans by rationally deciding to borrow more to allow for consumption smoothing, leading to higher debt levels. In addition, low-income students often receive grant-based aid (including federal Pell grants as well as institutional awards) which reduce the expected cost of college and reduce pressure to borrow. These sorts of differences across households raise the broader issue that even if increased borrowing makes sense on average, there can be considerable variation in realized borrowing, even among students with similar expected gains from collegiate attainment.

Uncertainty and Heterogeneity across Individuals

To this point, we have focused on the college investment and borrowing decisions on average; however, substantial variation in expected returns at the time of college entry for individuals may lead to different conclusions about the investment value of college and the associated level of borrowing. First, ultimate educational attainment varies considerably: some students will start but not complete college, while others will go on to complete graduate degrees that can pave the way to lucrative careers. Second, choice of occupations varies considerably, some with higher and some with lower average wages, among those students who achieve a given level of educational attainment. Third, substantial dispersion in wages exists even conditional on educational attainment and (broad) choice of occupation. In this section, we discuss these three factors, and the implications for the expected financial returns to college for a given student. As students make borrowing decisions, a central question is the extent to which they can accurately predict these determinants of future earnings. If students can accurately predict these determinants of future earnings, we would expect borrowing to vary substantially with these outcomes.

Collegiate Attainment

Only 55 percent of dependent students who anticipate completing a BA degree actually do so within six years of graduating high school, while more than one-third of them do not complete any postsecondary degree within six years. Similarly, more than half of dependent students who anticipate completing an associate's degree do not do so within six years of graduating high school (authors' tabulations, Beginning Postsecondary Study 2004:2009). Table 2 shows expected degree completion, realized degree completion, and the associated distribution of borrowing. One particularly negative outcome emerges: among students who anticipate completing a BA degree, 51.3 percent will end up with no degree and an average of $7,413 in student loans ($14,457 conditional on having borrowed at all).

To some degree, differences in educational outcomes across the set of college freshmen can be predicted by factors that are observable at the time of college entry. Not surprisingly, Bound, Lovenheim, and Turner (2010) show substantial differences in

Table 2
Expected Degree Completion, Realized Degree Completion, and Borrowing

	Expected attainment			
	No degree	*Certificate*	*AA*	*BA*
Distribution by expected attainment	3.6%	4.0%	13.0%	79.4%
Realized attainment				
No degree	66.2%	51.9%	62.0%	38.0%
Certificate	7.0%	31.9%	9.1%	2.1%
AA	4.9%	5.5%	21.5%	7.5%
BA	21.9%	10.8%	7.3%	52.4%
Percentage with student loans				
No degree	35.6%	37.0%	39.2%	51.3%
Certificate	22.0%	29.8%	47.9%	43.8%
AA	54.6%	35.1%	54.7%	55.6%
BA	66.3%	42.8%	65.4%	63.7%
Average student loans (unconditional)				
No degree	$4,475	$4,128	$4,222	$7,413
Certificate	$1,618	$2,788	$4,794	$5,113
AA	$7,651	$3,565	$8,544	$9,564
BA	$22,183	$9,658	$16,645	$15,562
Average student loans among borrowers				
No degree	$12,571	$11,160	$10,758	$14,457
Certificate	$7,367	$9,361	$10,008	$11,666
AA	$14,006	$10,149	$15,609	$17,194
BA	$33,480	$22,582	$25,465	$24,437

Source: Authors' tabulations from Beginning Postsecondary Survey (BPS) 2004:2009, including survey results in 2008–09 for students who entered any four-year college or public two-year college in 2003–2004.
Note: AA is Associate's degree; BA is Bachelor's degree.

degree completion rates conditional on student achievement. In addition, graduation rates and expected future earnings may differ among colleges and universities, perhaps because U.S. colleges and universities differ widely in available resources. Tabulations specific to this paper show that among students beginning at four-year colleges, private for-profit colleges have dramatically lower average graduation rates (16 percent) for dependent students than do public (63 percent) or private not-for-profit (68 percent) colleges. In addition, there is substantial variation in graduation rates within each college category, with more-selective colleges typically having higher graduation rates.[8]

[8] There is some debate in the literature about whether the economic benefits of attending a more-selective college can be explained entirely by selection, because more-promising students tend to attend more-selective colleges (for example, Hoxby 2001; Black, Daniel, and Smith 2005; Hoekstra 2009; Dale and Krueger 2011). But for the purpose of assessing the expected willingness to borrow, this debate is mostly immaterial—the question of interest to any particular student is "What is my expected financial gain

A student's computation of the expected financial return to entering college should incorporate the conditional probability of not completing college given all known factors, including that student's past achievement and the historical graduation rates for the college chosen. These adjustments would have more effect in reducing the expected value of attending higher education for students with lower achievement levels and especially for those attending colleges—such as private for-profit colleges—with very low documented graduation rates.

Choice of College Major and Career

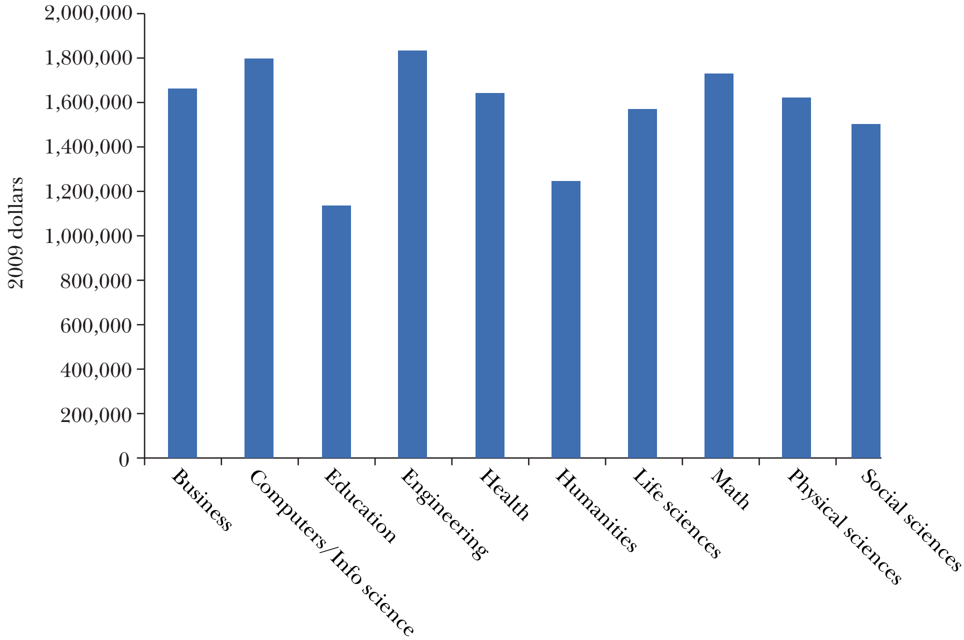
One widely cited story about a student struggling with an unusual amount of debt is the case of a 26 year-old graduate from New York University with $97,000 in loans referenced in a May 2010 *New York Times* story (Leiber 2010). With an interdisciplinary degree in religious and women's studies—which are fields of study with quite low expected earnings—one is left to wonder how this student's expectations about future earnings aligned with her borrowing decisions, both at the start of her college career and as she settled on her choice of major. Plainly, the student's prospects of paying the loan back are somewhat limited with a $22/hour job working for a photographer. Could this student have predicted the divergence between her earnings and her capacity to repay the loan? In practice, there are substantial differences in the expected lifetime earnings by choice of major.

Figure 4 shows the present discounted value of predicted lifetime earnings associated with different fields of specialization for men with exactly a BA. The estimates are based on a regression with the log of annual earnings as the dependent variable, and dummy variables for undergraduate major, post-baccalaureate degree attainment, job experience, race, and gender as explanatory variables. Not surprisingly, students who have chosen a technical field—in the broad categories of computer science, engineering, and math—tend to earn more than the average and more than those with education or humanities undergraduate concentrations. There is a substantial economics literature on the return to different undergraduate specializations including Paglin and Rufolo (1990), Grogger and Eide (1995), and dynamic models like Arcidiacono, Hotz, and Kang (2010). There are also more accessible publications available through public policy and career services sources (like Carnevale, Strohl, and Melton 2011), although it is not clear that students use this information when selecting a college major and choosing how much to borrow for college.

If students enter college with knowledge of their intended major, we would expect to see systematic differences in borrowing by field of study in relation to the expected earnings by field of study. Of course, some students enter college with no specific choice of major or career field in mind, while others may change their majors while enrolled in college, which in either case makes it difficult to take this factor into account in advance.

(or loss) given that I am attending college instead of taking a full-time job and not "What would be my expected financial gain (or loss) if I attend more-selective College *Y* rather than less-selective College *Z*?"

Figure 4
Expected Lifetime Earnings by Undergraduate Major, 2008



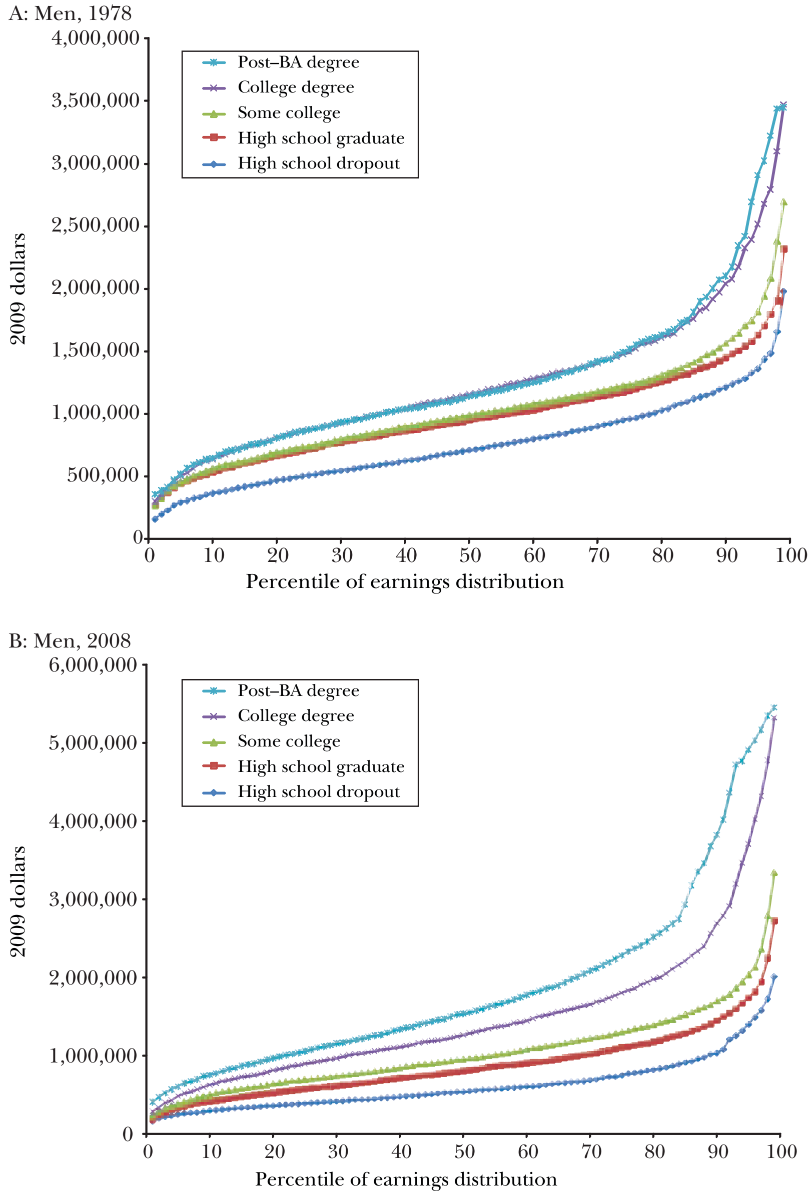
Source: Authors using data from the American Community Survey (2009).
Note: Based on regression of log annual earnings on dummy variables for undergraduate major, post-baccalaureate degree attainment, a quartic in experience, and indicators for race and gender using data from the American Community Survey (2009) with sample weights.

Collegiate Investment and the Increased Dispersion in Earnings and Attainment

As the average earnings of college graduates has increased, so too has the variance in earnings, and gains have been disproportionately concentrated among graduates with professional degrees and those with earnings outcomes in the top deciles (Acemoglu and Autor 2010; Lindley and Machin 2011). Annual differences in earnings among college graduates are magnified over the life course and, in turn, have a substantial impact on the expected return to a collegiate investment.

Figure 5 presents the distributions of lifetime earnings for different levels of postsecondary attainment for men in 1978, which is approximately the trough in the return to a college education, and 2008. (See the online Appendix available with this paper at ⟨http://e-jep.org⟩ for a similar figure for women, a group for whom participation in the labor market changed substantially during this time.) In both years, distribution of lifetime earnings for those with graduate degrees dominates the distribution for those with BA degrees, which in turn dominates the distribution for those completing some college, which in turn dominates the distribution for high school graduates. The difference in outcomes across these distributions widens markedly at the top part of the distribution beyond about the 80th percentile.

Figure 5
Distribution of Present Discounted Value of Career Earnings, Men



Source: Data are from the 1979 and 2009 annual files of the March CPS and are limited to full-time, full-year workers.
Notes: Percentiles of age-specific earnings profiles in each year are discounted to generate the expected value of lifetime earnings assuming a discount rate of 3 percent. More details of the calculations are presented in the online Appendix.

Figure 5 suggests two conspicuous changes from 1978 to 2008. First, differences in earnings between different postsecondary outcomes are more pronounced in 2008 than in 1978, with especially large gaps in 2008 between graduate degree recipients and BA degree recipients, and separately for BA degree recipients and those completing less than a BA degree. Secondly, within each education group, the difference between the median and the top of the distribution is much larger in 2008 than in 1978. To illustrate, the difference in expected lifetime earnings between a male college graduate at the 90th percentile and a male college graduate at the median rises from $963,149 in 1978 to $2,287,067 in 2008. For those who complete a graduate degree and find themselves in the top part of the distribution, the difference in earnings in 2008 relative to 1978 is extraordinary—on the order of $1.7 million over a lifetime for a man at the 90th percentile. For those who attend college and do not receive a degree, outcomes are notably stagnant, particularly in the lowest two-thirds of the distribution. Among men, those who have attended college but not received a BA degree are actually somewhat worse off over the life course in 2008 relative to 1978, while women in this situation are only modestly better off in 2008 than in 1978.

As we consider the increased observed variance in earnings within postsecondary outcomes, a key question is whether individuals are able to predict their position in the earnings distribution at the start of college and as they are making within-college borrowing decisions. If individuals have such information, we would expect borrowing to increase with an individual's place in the earnings distribution. Alternatively, the increase in the variance in earnings over time may reflect increased uncertainty about the economic outcomes associated with any educational trajectory.

While the relative importance of heterogeneity and uncertainty provide one framework for considering differences in collegiate investments and borrowing, it may be that student borrowing and investment decisions are also affected by imperfect information. If students systematically misperceive the likelihood of collegiate attainment or expected earnings, they may make "mistakes" in borrowing too much (or too little).

Implications for Borrowing and Collegiate Investments

How does the variation in the likelihood of completing a degree, choice of major, or where one will end up on the income distribution affect the decision to invest in a college education and, in turn, the decision to borrow for college? If individuals can make accurate predictions about whether they will complete college and what they would earn conditional on attaining a college degree, then most of the variation in lifetime earnings outcomes can be attributed to heterogeneity that is observable at the time of the decision—differences in individual aptitude or preparation, choice of college, and so forth. If, instead, individual characteristics that are observable at the time of college enrollment provide little information about future educational attainment and subsequent labor market outcomes, then an individual's best estimate of the financial return to enrolling in college, and how much to borrow, must be based on a probabilistic assessment of earnings, which may encompass a wide range of outcomes. In effect, are realized differences in earnings a result of heterogeneity or uncertainty?[9]

To illustrate the implications of these two different cases, consider hypothetical scenarios based on the correlation between a student's rank order in the distribution of career earnings for college graduates (assuming that this student attends and graduates from college) and that student's rank order in the distribution for high school graduates (assuming instead that the student does not go to college). At one extreme, an individual would have the same position in the rank order distribution of earnings at each degree level—so that someone at the 80th percentile of the high school distribution could expect to be at the 80th percentile of the collegiate distribution. At the other extreme, the correlation between an individual's position in the high school distribution and the college distribution is zero, which means that the best estimate of the outcome will be the earnings outcome for the person at the median of the college distribution. An intermediate case is the assumption of a correlation coefficient between postsecondary and college outcomes on the order of 0.75.[10]

How do these projections differ across the three decades from 1978 to 2008, given the appreciable gains at the very top of the collegiate wage distribution? Table 3 presents estimates under the three alternative assumptions of the high school–college correlation in rank ($\rho = 0, 0.75, 1$); the top panel shows the expected present value of net lifetime earnings of a college graduate and the bottom panel shows the expected differential between collegiate and high school earnings. Assuming perfect correlation between high school rank and college rank produces the distributions with the steepest upward trajectories—increasing earnings. To illustrate, a man at the 90th percentile of the high school, career-earnings distribution would be projected to have net collegiate, career earnings of \$1.8 million in 1978 (constant dollars) and \$2.3 million in 2008, while a student at the 10th percentile of the high school distribution would be projected to have career earnings of \$603,624 in 1978 and the slightly lower outcome of \$570,865 in 2008. As uncertainty increases, or the correlation coefficient decreases, projected career earnings "flatten" across the baseline distribution. With a weaker correlation, a greater share of the distribution

[9] Recent work in applied econometrics including Chen (2008) and Cunha, Heckman, and Navarro (2005) addresses the challenges of measuring the extent to which the potential dispersion of earnings is attributable to individual heterogeneity or uncertainty. In general, the problem of distinguishing heterogeneity from uncertainty is complicated by the absence of clear identification without very strong functional form assumptions. Chen (2008) attributes much of the greater wage inequality among college graduates than high school graduates to relatively larger effects of heterogeneity among individuals, though she estimates that about 80 percent of potential wage inequality among college graduates is attributable to uncertainty.

[10] In essence, we match the percentile of the high school distribution (*HS*) to a percentile in the college distribution as a conditional expectation which is a function of the correlation between *HS* and *C*, such that $E(C|HS) = (1-\gamma)\overline{HS} + \gamma HS$ where γ is the square of the correlation coefficient and $\overline{HS}$ is the average percentile (the median). When the correlation is 0.75, gamma is equal to 0.5625, and the expected rank in the college distribution is a weighted average of the median and the observed high school rank. Expected earnings are computed as a share-weighted combination of the earnings distributions for those at the different levels of collegiate attainment from less than a BA to graduate degrees.

Table 3
Projected Net Lifetime Earnings with a College Degree, and College–High School Differential, Alternative Assumptions, 1978 and 2008

Percentile of earnings distribution	1978			2008		
	$\rho = 1$	$\rho = .75$	$\rho = 0$	$\rho = 1$	$\rho = .75$	$\rho = 0$
Expected PDV collegiate earnings						
Women						
10	378,180	492,778	579,126	436,140	588,301	901,908
25	492,775	564,747	579,126	603,199	697,208	901,908
50	646,042	646,042	579,126	840,787	840,787	901,908
75	824,934	736,132	579,126	1,154,114	996,244	901,908
90	1,012,494	824,934	579,126	1,571,831	1,195,242	901,908
95	1,162,783	869,542	579,126	1,929,559	1,274,998	901,908
99	1,418,246	930,391	579,126	3,337,826	1,470,333	901,908
Men						
10	603,625	838,579	1,138,378	570,865	786,540	1,261,489
25	810,516	926,495	1,138,378	795,659	941,104	1,261,489
50	1,072,293	1,072,293	1,138,378	1,143,475	1,143,475	1,261,489
75	1,372,471	1,217,483	1,138,378	1,639,365	1,413,594	1,261,489
90	1,814,302	1,366,319	1,138,378	2,357,862	1,734,813	1,261,489
95	2,172,964	1,440,285	1,138,378	3,337,949	1,871,475	1,261,489
99	3,095,903	1,552,359	1,138,378	5,031,368	2,278,956	1,261,489

Percentile of high school wage distribution	1978			2008		
	$\rho = 1$	$\rho = .75$	$\rho = 0$	$\rho = 1$	$\rho = .75$	$\rho = 0$
Expected net returns (Coll-HS)						
Women						
10	134,813	249,411	335,759	203,784	355,945	669,552
25	167,899	239,871	254,250	305,930	399,938	604,639
50	221,041	221,041	154,125	439,393	439,393	500,514
75	273,940	185,138	28,132	644,108	486,238	391,902
90	296,916	109,356	–136,452	886,685	510,096	216,762
95	322,090	28,849	–261,567	1,108,659	454,099	81,008
99	292,183	–195,673	–546,937	2,327,401	459,908	–108,517
Men						
10	235,318	470,272	770,070	277,920	493,595	968,544
25	298,880	414,859	626,742	408,338	553,783	874,169
50	361,800	361,800	427,885	607,035	607,035	725,050
75	410,767	255,779	176,674	896,534	670,763	518,658
90	599,569	151,586	-76,355	1,326,780	703,732	230,408
95	813,152	80,473	–221,434	1,941,625	475,151	–134,834
99	1,116,979	–426,564	–840,546	3,024,145	271,733	–745,734

Source: Authors.

(and, at the extreme, the entire distribution) is expected to benefit from the rise in the return to collegiate attainment between 1978 and 2008, with this increase particularly large for women. In essence, with substantial uncertainty and rising benefits to college, more people would be expected to "give college a try," though as a result of this uncertainty, some college graduates would be expected to achieve smaller gains to college attainment than the values indicated in Table 3.

Yet this initial presentation assumes that high school graduates considering college essentially can uniformly expect to receive the earnings drawn from the distribution of all collegiate outcomes, including graduate attainment (where changes in returns have been the greatest in the last three decades). This estimate is surely an extreme upper bound for the students currently in the high school graduate pool.

Additional Factors

Two additional factors may have important implications for the financial costs and gains of enrolling in college: risk aversion and option value. Since enrolling in college can be viewed as a lottery with substantial probability of amassing debt but earning no degree, risk aversion would likely reduce the attractiveness of borrowing to enroll in college. At the same time, students can anticipate a flow of new information about costs (for example, time and effort required to complete a degree) and benefits (likely job placement and salaries) of college while enrolled. Since it is possible to drop out at any time, this flow of information induces an option value to initial college enrollment. Indeed, Stange (forthcoming) estimates that 14 percent of the (positive) expected return to college enrollment can be attributed to this option value (see also Stinebrickner and Stinebrickner 2009).

One important implication of the option value of enrolling in college is that even assuming optimal enrollment decisions by students based only on the financial implications of college (excluding, for example, the consumption value of attending college), we should still expect to see some students dropping out. George Stigler once commented, "If you never miss a plane, you're spending too much time at the airport." Similarly, if no one dropped out of college, we could likely conclude that more students should be enrolling.

Do Students Make Optimal Use of Loans in Financing College?

While it is too early to assess the extent to which early twenty-first century student borrowers as a group will face oppressive long-term burdens from their student debt, a look at student outcomes six years after college enrollment provides some indication of whether it is likely that the current generation is part of a "debt bubble." Table 4 presents total accumulated student borrowing six years after college entrance by type of first institution.

Table 4 also highlights the widespread variation in borrowing levels. Borrowing among students at the median is relatively modest: zero for students beginning at

Table 4
Borrowing Distribution after Six Years, by Degree Type and First Institution

	Type of institution of first enrollment			
	Public 4-year	*Private nonprofit 4-year*	*Private for-profit 4-year*	*Public 2-year*
All students beginning in 2004				
% Borrowing	61%	68%	89%	41%
Percentile of borrowers				
10th	$0	$0	$0	$0
25th	$0	$0	$6,376	$0
50th	$6,000	$11,500	$13,961	$0
75th	$19,000	$24,750	$28,863	$6,625
90th	$30,000	$40,000	$45,000	$18,000
Mean	**$11,706**	**$16,606**	**$19,726**	**$5,586**
BA recipients				
BA completion	61.5%	70.7%	14.8%	13%
% Borrowing	59%	66%	92%	69%
Percentile of borrowers				
10th	$0	$0	$12,000	$0
25th	$0	$0	$30,000	$0
50th	$7,500	$15,500	$45,000	$11,971
75th	$20,000	$27,000	$50,000	$23,265
90th	$32,405	$45,000	$100,000	$40,000
Mean	**$12,922**	**$18,700**	**$45,042**	**$15,960**

Source: Authors' tabulations based on the Beginning Postsecondary Survey 2004:2009.

community colleges, $6,000 for students at four-year public colleges, and $11,500 for students at private nonprofit colleges. Even at the 90th percentile, student borrowing does not exceed $40,000 outside of the for-profit sector. Examples of students who complete their undergraduate degree with more than $100,000 in debt are clearly rare: outside of the for-profit sector, less than 0.5 percent of students who received BA degrees within six years had accumulated more than $100,000 in student debt. The 90th percentile of degree recipients starting at for-profits have $100,000 in debt; so a nontrivial number of students at for-profits accumulate this much debt, but the situation is still far from the norm.

Leaving aside extreme cases, are student borrowing levels assumed by the majority of undergraduate students consistent with their capacity to repay these loans? There is little evidence to suggest that the average burden of loan repayment relative to income has increased in recent years. The most commonly referenced benchmark is that a repayment to gross income ratio of 8 percent, which is derived broadly from mortgage underwriting, is "manageable" while other analysis such as a 2003 GAO study set the benchmark at 10 percent. To put this in perspective, an individual with $20,000 in student loans could expect a monthly payment of about $212, assuming a ten-year repayment period. In order for this payment to accrue to

10 percent of income, the student would need an annual income of about $25,456, which is certainly within the range of expected early-career wages for college graduates. Overall, the mean ratio of student loan payments to income among borrowers has held steady at between 9 and 11 percent, even as loan levels have increased over time (Baum and Schwarz 2006; Baum and O'Malley 2003). Among student borrowers in repayment six years after initial enrollment, the mean ratio of monthly payments to income is 10.5 percent (author's tabulations from the Beginning Postsecondary Study 2004:2009).

Table 4 also highlights differential levels of borrowing by first institution of attendance. In particular, the borrowing behavior among students beginning at for-profit institutions is distinctly higher at all levels of credit attainment than among students at other types of postsecondary institutions. These systematic differences in borrowing translate predictably into differences in default rates by first institution of attendance. Data from the Department of Education on the Official Cohort Default Rates for Schools (available at ⟨http://www2.ed.gov/offices/OSFAP/defaultmanagement/cdr.html#table⟩) shows two-year cohort default rates rising from 6.7 to 8.8 percent between 2007 and 2009. At for-profit institutions, default rates are appreciably greater, reaching 15 percent over two years and 24.9 percent over three years. Student characteristics are insufficient to account for these high default rates in the for-profit sector (as discussed by Deming, Goldin, and Katz in this symposium), which suggests that students choosing to attend these institutions may be systematically borrowing too much.

Student Loans and Financial Portfolios

Even when college is a "good investment" in a net present value sense, students may finance it badly. Do students borrow the "right" amount for college? Do they borrow from the lowest cost of capital? Even if some students may borrow "too much" for college, other students may make the opposite mistake, "underborrowing" by insufficient use of student loans in financing college.

Cadena and Keys (2010) estimate that one in six full-time students at four-year institutions who are eligible for student loans do not take up such loans—thus forgoing the subsidy.[11] The most obvious explanations for this behavior are that some students are deterred by the complexity of the FAFSA form (Dynarski and

[11] A growing number of community college students do not have access to federal Stafford loans. For students entering college in 1992–93, less than 3 percent of community college students did not have access to Stafford loans (calculated from National Longitudinal Study of 1988). For community college students entering college in 2004–2005, about 11 percent of students did not have access to loans (Educational Longitudinal Study of 2002). The Project on Student Debt report found similar numbers, and in their April 2011 report, they calculated that 9 percent of community college students do not have access to Stafford loans. One explanation for why a community college might not offer loans is that if an institution has a default rate over 25 percent for three consecutive years or if a community college has a default rate of 40 percent in one year, the institution will lose access to Title IV funds (including Pell grants). But few community colleges are near the default thresholds.

Scott-Clayton 2006) or that students rationally avoid student loans as a self-control device (Cadena and Keys 2010). Another possible sign of the underuse of student loans is that a number of students are carrying more-expensive credit card debt when they could instead be borrowing through student loans. Among students who entered college in 2004, 25.5 percent of those who were still enrolled in 2006 and 37.7 percent of those who were still enrolled in 2009 reported that they had credit card debt. But between one-third and one-half of these students (45.6 percent of students with credit card debt in 2006 and 38.5 percent of students with credit card debt in 2009) had not borrowed from the Stafford loan program. Carrying credit card debt without maximizing Stafford borrowing burdens students with unnecessarily inflated interest rates—a choice that can interfere with a student's ability to finish a degree: some years back, a school administrator, John Simpson at Indiana University, said: "[W]e lose more students to credit card debt than academic failure" (Rubin 1998). Along similar lines, about half of the students who are working more than 20 hours per week while attending a public or private nonprofit four-year college have no Stafford loans at all (authors tabulations from Beginning Postsecondary Study 2004:2009). But since there is some evidence that part-time work reduces academic performance and the likelihood of attaining a degree (Stinebrickner and Stinebrickner 2003), it might be optimal for some of these students to work fewer hours and use Stafford Loans to substitute in the short- and medium-term for lost wages.

Conclusions and Further Thoughts

Enrolling in college is likely the first major capital investment that young people will make. For many students, it will be their first encounter with a formal loan. From a financial perspective, enrolling in college is equivalent to signing up for a lottery with large expected gains—indeed, the figures presented here suggest that college is, on average, a better investment today than it was a generation ago—but it is also a lottery with significant probabilities of both larger positive, and smaller or even negative, returns.

The natural advice for a high school graduate contemplating the economic consequences of investing in college is to estimate the probabilities of the long-term outcomes as precisely as possible. In particular, a student needs to focus on the probability of degree completion, the earnings differences associated with different levels of degree completion, and the choice of a field of study. Although self-knowledge is difficult, students can look at their own observed traits, and then at how students with similar traits have fared at the school they are planning to attend. For example, those who begin their studies at community colleges and for-profit colleges have particularly low college completion rates and are unlikely to realize substantial earnings gains associated with degree completion. For students at for-profit institutions, the consequences of weak outcomes are compounded by high levels of borrowing; not surprisingly, these students are unusually likely to default on loans. Perhaps the hardest risk to estimate involves the substantial and increasing variation in realized earnings within different levels of postsecondary attainment: for students who end up in the bottom part of the wage distribution (given attainment in college), debt levels are likely higher than their earnings would justify.

The claim that student borrowing is "too high" across the board can—with the possible exception of for-profit colleges—clearly be rejected. Indeed, media coverage proclaiming a "student loan bubble" or a "crisis in student borrowing" even runs the risk of inhibiting sound and rational use of credit markets to finance worthwhile investments in collegiate attainment. McPherson and Baum (2011) note that one form of cognitive bias impacting collegiate investments is attaching too much significance to extreme examples, like the few instances of undergraduate students burdened with more than $100,000 in debt with poor job prospects. Even if macroeconomic shocks were to erode the higher education earnings premium to levels not seen in three decades, collegiate attainment would remain a good investment for many potential students. Given the relatively slow rate of growth in the supply of college graduates in recent decades and modest projections for further increases in the coming decades, it is highly unlikely that the economy will experience a demand shock that will have a substantial adverse impact on the wages of college graduates.

The observation that college is a good investment for most young people still leaves a number of significant and unanswered research questions about how students make decisions about collegiate attainment and student borrowing. In the context of this paper, an especially important question would be to assess more carefully what verifiable characteristics students could observe about their own skills and attributes at the time of college entry which in turn would affect their outcomes both in higher education and in the workplace later in life. Student decisions about whether to enroll in college, where to enroll in college, what to study in college, and how to finance college are complex and highly dependent on individual circumstances. While some uncertainty will inevitably remain about the decision of whether and how to invest in higher education, it seems clear that a substantial number of students could benefit from more-tailored and individualized advice than they have been receiving.

■ *We thank the JEP Editors and Assistant Editor for their patience and guidance and also Adrew Barr and Erin Dunlop for research assistance.*

References

Acemoglu, Daron, and David Autor. 2010. "Skills, Tasks and Technologies: Implications for Employment and Earnings." Chap. 12 in *Handbook of Labor Economics,* Volume 4B, edited by Orley Ashenfelter and David Card, Amsterdam: Elsevier, North Holland.

Arcidiacono, Peter; V. Joseph Hotz, and Songman Kang. 2010. "Modeling College Major

Choices using Elicited Measures of Expectations and Counterfactuals." NBER Working Paper 15729.

Avery, Christopher, and Thomas J. Kane. 2004. "Student Perceptions of College Opportunities. The Boston COACH Program." Chap. 9 in *College Choices: The Economics of Where to Go, When to Go, and How to Pay For It,* edited by Caroline M. Hoxby. University of Chicago Press.

Baum, Sandra, and Marie O'Malley. 2003. "College on Credit: How Borrowers Perceive Their Education Debt." *Journal of Student Financial Aid* 33(3): 7–21.

Baum, Sandy, and Saul Schwartz. 2006. "How Much Debt is Too Much? Defining Benchmarks for Manageable Student Debt." Commissioned by the College Board and the Project on Student Debt. http://professionals.collegeboard.com/prof download/pdf/06-0869.DebtPpr060420.pdf.

Black, Dan, Kermit Daniel, and Jeffrey Smith. 2005. "College Quality and Wages in the United States." *German Economic Review* 6(3): 415–43.

Black, Dan, and Jeffrey Smith. 2006. "Evaluating the Returns to College Quality with Multiple Proxies for Quality." *Journal of Labor Economics* 24(3): 701–28.

Bound, John, Michael Lovenheim, and Sarah Turner. 2010. "Why Have College Completion Rates Declined? An Analysis of Changing Student Preparation and Collegiate Resources. *American Economic Journal: Applied Economics* 2(3): 129–57.

Burdman, P. 2005. "The Student Debt Dilemma: Debt Aversion as a Barrier to College Access." Project on Student Debt, Institute for College Access and Success, Inc.

Cadena, Brian C., and Benjamin J. Keys. 2010. "Can Self-Control Explain Avoiding Free Money? Evidence from Interest-Free Loans," December. http://spot.colorado.edu/~cadenab/Research_files/cadenakeys_loans_resub.pdf.

Carnevale, Anthony, Jeff Strohl, and Michell Melton. 2011. *What's it Worth? The Economic Value of College Majors.* Georgetown University, Center on Education and the Workforce.

Chen, Stacey. 2008. "Estimating the Variance of Wages in the Presence of Selection and Unobserved Heterogeneity." *Review of Economics and Statistics* 90(2): 275–89.

College Board, Advocacy and Policy Center. 2010a. "Trends in Student Aid: 2010." http://advocacy.collegeboard.org/sites/default/files/2010_Student_Aid_Final_Web.pdf.

College Board, Advocacy and Policy Center. 2010b. "Trends in College Pricing." http://advocacy.collegeboard.org/sites/default/files/2010_Trends_College_Pricing_Final_Web.pdf.

Cunha, Flavio, James J. Heckman, and Salvador Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings." *Oxford Economic Papers* 57(2): 191–261.

Dale, Stacy, and Alan Krueger. 2011. "Estimating the Return to College Selectivity over the Career Using Administrative Earning Data." http://www.mathematica-mpr.com/publications/PDFs/education/returntocollege.pdf.

Day, Jennifer Cheeseman, and Eric Newburger. 2002. "The Big Payoff: Educational Attainment and Synthetic Estimates of Work-Life Earnings." U.S. Census Bureau, Current Population Reports, P23-210, July. http://www.census.gov/prod/2002pubs/p23-210.pdf.

Dickler, Jessica. 2010. "Boomerang: 85% of College Graduates Move Back Home." CNNMoney, November 15. http://money.cnn.com/2010/10/14/pf/boomerang_kids_move_home/index.htm.

Dominitz Jeffrey, and Charles Manski. 1996. "Eliciting Student Expectations of the Returns to Schooling." *Journal of Human Resources* 31(1): 1–26.

Dynarski, Judith, and Judith Scott-Clayton. 2006. "The Cost of Complexity in Federal Student Aid: Lessons from Optimal Tax Theory and Behavioral Economics." NBER Working Paper 12227.

Ellwood, David. 2001. "The Sputtering Labor Force of the 21st Century. Can Social Policy Help?" Chap. 9 in *The Roaring Nineties: Can Full Employment Be Sustained?* Edited by Alan Krueger and Robert Solow. New York: Russell Sage Foundation.

Freeman, Richard. 1976. *The Overeducated American.* New York: Academic Press.

Friedman, Milton. 1962. *Capitalism and Freedom.* University of Chicago Press.

Gicheva, Dora. 2011. "Does the Student-Loan Burden Weigh into the Decision to Start a Family?" March. http://www.uncg.edu/bae/people/gicheva/Student_loans_marriageMarch11.pdf.

Goldin, Claudia, and Lawrence Katz. 2008. *The Race between Education and Technology*. Cambridge, MA: Harvard University Press.

Grogger, Jeff, and Eric Eide. 1995. "Changes in College Skills and the Rise in the College Wage Premium." *Journal of Human Resources* 30(2): 280–310.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics* 91(4): 717–24.

Hoxby, Caroline. 2001. "The Return to Attending a More Selective College: 1960 to Present." Chap. 2 in *Forum Futures Exploring the Future of Higher Education, 2000 Papers,* edited by Maureen Devlin and Joel Meyerson. San Francisco: Jossey-Bass.

Institute of Education Sciences. 2011. *Digest of Education Statistics: 2010* http://nces.ed.gov/programs/digest/d10/index.asp. National Center for Education Statistics, U.S. Department of Education.

Kahn, Lisa. 2010. "The Long-Term Labor Market Consequences of Graduating from College in a Bad Economy." *Labour Economics* 17(2): 303–316.

Kamenetz, Anya. 2006. *Generation Debt.* Riverhead Books.

King, Jacqueline E. 2002. *Crucial Choices: How Students' Financial Decisions Affect Their Academic Success.* Washington, DC: American Council on Education, Center for Policy Analysis.

Krueger, Alan B., and William G. Bowen. 1993. "Policy Watch: Income-Contingent College Loans." *Journal of Economic Perspectives* 7(3): 193–201.

Leiber, Ron. 2010. "Placing the Blame as Students Are Buried in Debt." New *York Times,* May 28. http://www.nytimes.com/2010/05/29/your-money/student-loans/29money.html.

Lindley, Joanne, and Stephen Machin. 2011. "Rising Wage Inequality and Postgraduate Education." CEP Discussion Papers dp1075, Center for Economic Performance, London School of Economics.

Lochner, Lance J., and Alexander Monge-Naranjo. 2011. "The Nature of Credit Constraints and Human Capital." *American Economic Review* 101(6): 2487–2529.

Lochner, Lance, and Alexander Monge-Naranjo. Forthcoming. "Credit Constraints in Education." *Annual Review of Economics.*

Long, Bridget. 2004. "The Role of Perceptions and Information in College Access: An Exploratory Review of the Literature and Possible Data Sources." July. Report commissioned by TERI, Boston, MA.

Manski, Charles. 1993. "Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?" In *Studies of Supply and Demand in Higher Education,* edited by Charles Clotfelter and Michael Rothschild, 43–57. Chicago: University of Chicago Press.

Mazzeo, Christopher. 2007. "Private Lending and Student Borrowing: A Primer." In *Footing the Tuition Bill,* edited by Frederick M. Hess. Washington, DC: AEI Press.

McPherson, Michael, and Sandra Baum. 2011. "Get Smart about College." *Wall Street Journal,* September 19.

Oreopoulos, Philip, and Kjell G. Salvanes. 2011. "Priceless: The Nonpecuniary Benefits of Schooling." *Journal of Economic Perspectives* 25(1): 159–184.

Paglin, Morton, and Anthony M. Rufolo. 1990. "Heterogeneous Human Capital, Occupational Choice, and Male–Female Earnings Differences." *Journal of Labor Economics* 8(1): 123–44.

Rothstein, Jesse, and Cecilia Rouse. 2011. "Constrained after College: Student Loans and Early Career Occupational Choices." *Journal of Public Economics* 95(1–2): 149–63.

Rouse, Cecelia Elena. 2004. "Low-Income Students and College Attendance: An Exploration of Income Expectations." *Social Science Quarterly* 85(5): 1299–1317.

Rubin, Bonnie Miller. 1998. "College Students Charge Right into Valley of Debt," *Chicago Tribune,* August 16. http://articles.chicagotribune.com/1998-08-16/news/9808160239_1_credit-history-college-students-credit-card.

Schumpeter Blog. 2011. "Higher Education: The Next Bubble?" *Economist,* April 13. http://www.economist.com/blogs/schumpeter/2011/04/higher_education.

Smith, James P., and Finis Welch. 1978. "The Overeducated American: A Review Article." RAND Working Paper No. P-6253. http://www.rand.org/pubs/papers/P6253.html.

Somers, Patricia, Shawn Woodhouse, and Jim Cofer. 2004. "Pushing the Boulder Uphill: The Persistence of First-Generation College Students." *NASPA Journal* 41(3): 418–35.

St. John, Edward P. 2004. "The Impact of Information and Student Aid on Persistence: A Review of Research and Discussion of Experiments." October. Report commissioned by TERI, Boston, MA. http://www.teri.org/pdf/research-studies/ReseachReport_StJohn.pdf.

Stange, Kevin. Forthcoming. "An Empirical Examination of the Option Value of College Enrollment." *American Economic Journal: Applied Economics.*

Steven Ruggles, J. Trent Alexander, Katie Genadek, Ronald Goeken, Matthew B. Schroeder, and Matthew Sobek. 2010. Integrated Public Use Microdata Series: Version 5.0 (Machine-readable database). University of Minnesota.

Stinebrickner, Ralph, and Todd R. Stinebrickner. 2003. "Working during School and Academic Performance." *Journal of Labor Economics* 21(2): 449–72.

Stinebrickner, Todd R., and Ralph Stinebrickner. 2009. "Learning about Academic Ability and the College Drop-Out Decision." NBER Working paper 14810.

Surowiecki, James. 2011. "Debt by Degrees." *New Yorker,* November 21. http://www.newyorker.com/talk/financial/2011/11/21/111121ta_talk_surowiecki.

U.S. Department of Education. Nd. "Official Cohort Default Rates for Schools" Database.

Available at: http://www2.ed.gov/offices/OSFAP/defaultmanagement/cdr.html#table.

Warwick, Jacquelyn, and Phylis Mansfield. 2000. "Credit Card Consumers: College Students' Knowledge and Attitude." *Journal of Consumer Marketing* 17(7): 617–26.

Wei, Christina Chang. 2010. *Trends in Undergraduate Stafford Loan Borrowing: 1989–90 to 2007–08*. National Center for Education Statistics Web Tables. Washington, DC: U.S. Department of Education. http://nces.ed.gov/pubs2010/2010183.pdf.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 193–216

American Higher Education in Transition†

Ronald G. Ehrenberg

American higher education is in transition along many dimensions: tuition levels, faculty composition, expenditure allocation, pedagogy, technology, and more.

During the last three decades, at private four-year academic institutions, undergraduate tuition levels increased each year on average by 3.5 percent more than the rate of inflation. The comparable increases for public four-year and public two-year institutions were 5.1 percent and 3.5 percent, respectively (Baum and Ma 2011, figure 4). Tuition increases in private higher education have been associated over this period with increased real expenditures per student. In public higher education, as I detail below, at best, tuition increases have helped to compensate for reductions in state support (Desrouchers, Lenihan, and Wellman 2010).

The forces that cause private and public tuitions to increase at rates that exceed the rate of inflation have been extensively discussed in Ehrenberg (2002, 2006, 2007, 2010) and Archibald and Feldman (2011). They include the aspirations of academic institutions to be the very best they can be in every dimension of their activity. Also important are student and parent perceptions that where one goes to college is almost as important as whether one goes to college and the belief that higher-priced selective private institutions confer unique educational and economic advantages on their students; this leads higher-priced, selective private institutions to have long lines of applicants and only limited market forces to limit their tuition

■ *Ronald G. Ehrenberg is the Irving M. Ives Professor of Industrial and Labor Relations and Economics, and Director of the Cornell Higher Education Research Institute (CHERI), both at Cornell University, Ithaca, New York. He is also a Research Associate, National Bureau of Economic Research, Cambridge, Massachusetts. His e-mail address is ⟨rge2@cornell.edu⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.1.193.
doi=10.1257/jep.26.1.193

increases, which in turn provides cover for less-selective institutions to raise their tuition levels.[1] Higher education is also driven by published rankings, such as those of *U.S. News and World Report*, which are based partially on institutions' expenditures per student. Finally, the growth of technology can lead to improvements in the quality of higher education but often comes at a high cost. For public institutions, add to these pressures the cutbacks in state support.

Even as undergraduate tuition levels and spending per student are increasing, the nature of faculty positions has changed dramatically during the last 30 to 40 years. The percentage of faculty nationwide that is full-time has declined from almost 80 percent since 1970, to 51.3 percent in 2007, and the vast majority of part-time faculty members do not have Ph.D.s (Snyder and Dillow 2010, tables 249, 253). The percentage of full-time faculty not on tenure track has more than doubled between 1975 and 2007, increasing from 18.6 percent to 37.2 percent (AAUP Fact Sheet, n.d.). Of course, this change raises the question of whether, or how much, different types of undergraduates benefit from being taught by full-time tenured or tenure-track faculty.

Part of the reason for a rise in tuition at the same time as what appears to be a decline in spending on faculty is that the *tuition discount rate*—the share of each tuition dollar that institutions returned to their undergraduate students in the form of need-based or merit grant aid—increased substantially at private four-year institutions. For example, the average tuition discount rate for first-time, full-time, first-year students at private four-year institutions reached 42 percent in fall 2008; in fall 1990, the comparable figure was 26.7 percent (National Association of College and University Business Officers 2009, 2010). In short, much of the increase in tuition revenues at private colleges and universities has been plowed back into undergraduate aid; at all but a handful of the very wealthiest private institutions, the vast majority of undergraduate financial aid dollars come from tuition revenue.[2] The wealthiest and most selective private institutions of higher education dramatically increased the generosity of their financial aid policies for several reasons: relatively small fractions of their students were coming from lower-income and lower-middle-income families (Supiano and Fuller 2011), and the institutions wanted to attract

[1] That selective institutions provide students with unique advantages is disputed, with most studies, including Brewer, Eide, and Ehrenberg (1999), finding it to be true, while two other studies, Dale and Kruger (2002, 2011), offer contrary evidence.

[2] A different, but important, question is how the *net tuition cost* paid by the average student has changed over time. In addition to institutional grant aid, net tuition calculations adjust posted tuition rates for federal, state, and other private grant aid and for tax credits for educational expenses. The College Board reports that while average tuition levels at public and private not-for-profit four-year institutions grew by average annual rates of 7.0 and 5.3 percent, respectively, during the 1990–91 to 2011–12 period, *net tuition* at the two types of institutions grew at lower annual rates of 4.1 and 3.4 percent, respectively, during the period. Average tuition levels grew at average annual rate of 0.6 percent per year at public two-year colleges during the period, but net tuition actually declined at them, largely due to increases in the generosity of the federal Pell Grant program (unpublished data from the College Board provided by Sandy Baum). For comparison purposes, the average annual rate of increase in the Consumer Price Index during the 1990 to 2010 period was about 2.7 percent.

these students; a combination of rapid growth rates in their endowments during much of the period and relatively low endowment spending rates led to pressure from the U.S. Congress for them to increase endowment spending on financial aid; and, after the financial collapse in 2008, the decline in family incomes and asset levels meant dramatic increases in the financial need of their applicants. Other less-selective private institutions, which face highly salient competition from lower-priced public institutions, also faced a dramatic need to increase grant aid and offer tuition discounts both to fill all their seats and to achieve desired class composition in terms of student selectivity and other characteristics.[3]

In public higher education, tuition increases in recent decades have barely offset a long-run decline in state appropriations per full-time equivalent student. State appropriations per full-time equivalent student at public higher educational institutions averaged $6,454 in fiscal year 2010; at its peak in fiscal year 1987, the comparable number (in constant dollars) was $7,993 (State Higher Education Executive Officers 2011, figure 3), translating into a decline of 19 percent over the period. Even if one leaves out the "Great Recession," real state appropriations per full-time equivalent student were still lower in fiscal year 2008 than they were 20 years earlier. Overall, the sum of net tuition revenue and state appropriations per full-time equivalent student at the publics was roughly the same in real terms in fiscal year 2010 as it was in fiscal year 1987.

In addition, academic institutions have changed how they allocate their resources. The share of institutional expenditures going to faculty salaries and benefits in both public and private institutions has fallen relative to the share going to nonfaculty uses like student services, academic support, and institutional support (Desrochers, Lenihan, and Wellman 2010). This change has been accompanied by changing modes of instruction, together with different uses of technology—and in a number of schools by charging differential tuition across students.

This paper discusses these changes in faculty composition, expenditure allocation, pedagogy, technology, and differential tuition, how they are distributed across higher education sectors, and their implications. I conclude with some speculations about the future of American education.

The Changing Nature of the Faculty

The composition of the faculty in institutions of higher education has evolved in two ways: Ph.D.s have become more widespread among the full-time faculty across all types of institutions, but there has been a move away from full-time and tenure-track jobs.

[3] While tuition levels rose in percentage terms by more at the four-year publics than they did at the four-year privates during the period, because tuition levels were so much lower at the publics at the start of the period, dollar increases in tuition were much larger at the privates, and the difference between public and private tuition levels (in real terms) increased during the period.

Table 1
Changing Faculty Types

	Full-time, non-tenure-track faculty as a percentage of all full-time faculty (A)			*Part-time faculty as a percentage of all faculty* (B)		
Category (sample size)	*1995*	*2001*	*2007*	*1995*	*2001*	*2007*
Associate's						
Public (899)	38.4	39.4	43.1	64.7	67.0	68.9
Private not-for-profit (51)	74.3	75.4	82.5	52.3	50.4	56.1
Private for-profit (101)	98.7	90.0	97.7	49.0	51.0	57.7
Bachelor's						
Public (139)	17.1	22.9	23.4	39.6	42.2	43.7
Private not-for-profit (497)	22.2	26.9	30.8	33.1	37.4	41.7
Private for-profit (33)	79.6	91.9	90.6	57.9	64.9	78.6
Master's						
Public (261)	12.7	17.6	20.6	29.3	37.0	40.3
Private not-for-profit (332)	25.1	28.6	33.6	50.8	53.3	59.5
Private for-profit (17)	71.6	85.2	93.7	62.2	70.8	89.7
Doctoral						
Public doctoral (166)	24.4	32.1	35.2	19.7	22.5	24.0
Private not-for-profit (106)	18.2	35.4	46.2	32.2	34.9	31.7
Private for-profit (4)	98.3	100.0	100.0	44.8	72.3	84.0

Source: Author's calculations based on data for 2,606 institutions that reported information to the Integrated Postsecondary Education Data System (IPEDS) *Fall Staff Surveys* in all of the years.
Note: In this table, and several others that follow, institutions classified as "associate's" typically offer two-year degrees as the highest degree; those classified as "bachelor's" offer primarily bachelor's degrees; those classified as "master's" typically offer undergraduate and master's degrees; and those classified as "doctoral" typically offer a wide range of undergraduate and graduate degrees including doctoral degrees.

On the spread of Ph.D.s, the best historical data is collected annually by the American Mathematical Society (and is available at ⟨http://www.ams.org/profession/data/annual-survey/annual-survey⟩). Between 1967 and 2009, the share of full-time mathematics faculty with a Ph.D. remained constant at about 90 percent at departments that offered doctoral degrees, but rose from 40 to 80 percent at those whose highest degree offered was a master's degree and from 30 to 70 percent at departments whose highest degree offered was a bachelor's degree, with most of the increase in the latter two types of institutions occurring by the mid-1980s (Ehrenberg 2011, figure 4.1). Assuming that mathematics was typical of many other academic disciplines, a growing supply of Ph.D.s allowed the bachelor's and master's institutions to increase the shares of their full-time faculty members with Ph.D.s.

Columns A of Table 1 present information on the percentages of full-time faculty members that are not on tenure tracks, by institutional type, for 1995, 2001, and 2007. In this table, and several others that follow, institutions classified as "associate's" typically offer two-year degrees as the highest degree; those classified as "bachelor's" offer primarily bachelor's degrees; those classified as "master's" typically offer undergraduate and master's degrees; and those classified as "doctoral" typically offer a wide range of undergraduate and graduate degrees including doctoral degrees. The data are for a set of 2,606 institutions that reported information to the Integrated Postsecondary Education Data System (IPEDS) *Fall Staff Surveys* in all of the years.

During the period, the percentages of full-time faculty members that were not on tenure tracks increased at all categories of institutions, with the largest absolute increase occurring at private not-for-profit doctoral institutions. As the research intensity of doctoral institutions increased over time, more of the undergraduate instruction at these institutions is being undertaken by full-time, non-tenure-track faculty. I will discuss this pattern further below. The percentages of full-time faculty members not on tenure tracks are very high at all categories of for-profit institutions.

Columns B of Table 1 present similar information on the percentages of all faculty members who are part-time. The part-time percentage grew at all categories of institutions, save for the private, not-for-profit doctoral institutions. In many categories the growth was relatively modest over the time frame shown here, with the greatest growth occurring in the growing for-profit higher education sector. The vast majority of part-time faculty do not have doctoral degrees (Ehrenberg 2011, table 4.4).

Data on the changes that have occurred specifically in departments of economics are more limited. The American Economic Association's annual survey of economics departments collects information on faculty types and data for a matched sample of 59 institutions offering Ph.D.s and 86 institutions where bachelor's degrees are the highest offered. Data for academic years 1998–99 and 2008–2009 appear in Scott and Siegfried (2009). During this ten-year period, the percentages of full-time faculty that were not on tenure tracks in the AEA sample increased from 4.3 to 8.7 percent in the economics departments of Ph.D. institutions, and from 7.5 to 13.8 percent in economics departments of the bachelor's institutions; the percentages of faculty that were part-time increased at the same institutions from 3.9 to 7.9 and from 6.5 to 11.9 percent, respectively (Scott and Siegfried, table 5, panel C).

To confirm these results, which after all are based on a limited number of institutions, I put a couple of research assistants to work in February 2011 looking at the web pages of the faculty employed at institutions ranked by *U.S. News & World Reports:* in particular, the top 83 ranked Ph.D. programs in economics, the economics departments at the top 189 national liberal arts colleges, and the economics departments at the top 107 regional master's institutions.[4] They calculated the number of full-time faculty members that are tenured or on tenure track, the number of full-time faculty that are not on tenure track, and the number of full-time faculty that are visitors at each institution. In these calculations, instructors, lecturers, senior lecturers, clinical professors, professors of practice, and visiting

[4] Some of the departments at master's institutions are departments of "economics and . . ." In these cases, wherever possible, the tabulations were limited to faculty who were teaching economics.

Table 2
Full-Time Faculty in Economics Departments that Are Non-Tenure-Track or Visitors with Professorial Titles
(for top Ph.D. programs in economics, and top national liberal arts colleges and regional master's institutions as defined by U.S. News & World Report*)*

	Non-tenure-track		*Visitors with professorial titles*	
	Mean percentage	*(Mean percentages weighted by faculty size)*	*Mean percentage*	*(Mean percentages weighted by faculty size)*
Top 83 Ph.D. programs in economics	13.8	(15.0)	9.2	(9.5)
a) Top 25	13.1	(13.9)	15.9	(16.0)
b) Rank 27–50	15.4	(16.6)	5.9	(6.1)
c) Rank 54–83	13.0	(18.7)	6.4	(5.2)
Economics departments at the top 189 national liberal arts colleges	6.0	(6.5)	4.8	(6.6)
a) Top 50	6.6	(7.3)	11.6	(10.3)
b) Rank 51–100	5.7	(5.1)	3.8	(4.4)
c) Rank 101–189	5.8	(6.3)	1.0	(0.8)
Economics departments at the top 107 regional master's institutions	8.3	(11.8)	3.0	(3.6)

Source: Author's calculations from faculty data on departmental web pages in February 2011.
Notes: Faculty classified as non-tenure-track include lecturers, instructors, visiting lecturers, visiting instructors and faculty with titles such as professor of practice or clinical professor. Percentage "Non-tenure-track" are: Non-tenure-track faculty / (Tenured or tenure-track faculty + Non-tenure-track faculty). Percentage "Visitors with professorial titles" are: Visitors with professorial titles / (Tenured or tenure-track faculty + Non-tenure-track faculty).

instructors were counted as "not on tenure track." A separate tabulation looked at faculty with visiting professorial titles, because, especially at the major doctoral universities and selective liberal arts colleges, visitors may be tenured or on tenure tracks at other institutions. The research assistants then summarized the numbers over all of the departments in a category and computed the means (weighted by faculty size) across departments of the percentages of full-time faculty that are not on tenure track, excluding visitors other than visiting instructors. They also calculated the percentages of all full-time faculty that are visitors with professorial rank. These percentages appear in Table 2.[5]

The mean percentage of full-time economics faculty (excluding visitors with professorial ranks) that are *not* on tenure tracks is 13.8 percent for the top 83 Ph.D. programs, while the weighted (by faculty size) mean is 15.0 percent; both of these measures are higher than the comparable implied percentage found by Scott

[5] These calculations may understate the percentage of non-tenure-track faculty because some faculty with professorial ranks may not be on tenure tracks in these departments. Departmental web pages did not uniformly list all part-time faculty members so we could not tabulate information for this group.

and Siegfried (2009). The mean percentage of full-time visiting faculty in these departments is around 9 percent. Visiting professors make up a much greater share (around 16 percent) of the faculty at top 25 ranked economics departments, probably because tenured and tenure-track faculty from other leading departments often visit for research purposes.

The percentage of full-time economics faculty that are not on tenure tracks at top liberal arts colleges is around 6 percent; somewhat lower than Scott and Siegfried (2009) found. Visiting faculty members are much more prevalent in economics departments at the top 50 liberal arts colleges than they are at the other national liberal arts colleges. Finally, the mean percentage of full-time economics faculty members that are not on tenure tracks at top regional master's institutions is 8.3 percent and the weighted mean is 11.8 percent. Visiting professors are scarcer at these master's institutions relative to the other categories of institutions.

A final source of data on economics faculty comes from the annual reports of the Committee on the Status of Women in the Economics Profession (CSWEP). These reports provide data for a larger sample of Ph.D.-granting departments and liberal arts colleges than the AEA data, because of the persistence of CSWEP members in making contacts at each department. For example, the 2010 CSWEP report was based on data from 121 Ph.D.-granting institutions and 97 liberal arts institutions. The CSWEP data indicate that the percentage of full-time faculty that were not on tenure tracks at economics departments rose from 10.8 to 20.0 at the Ph.D. institutions and from 15.0 to 16.4 at the liberal arts institutions between 2005 and 2010.[6] Some care must be used in interpreting these numbers because the responding institutions vary between the two years and the CSWEP data do not separate out visiting faculty and other non-tenure-track faculty. But they do confirm that the usage of full-time, non-tenure-track faculty has been increasing at the doctoral universities. In these data, 33 percent of the non-tenure-track faculty in economics were female at the Ph.D. institutions in 2010; the comparable female share of tenure-track assistant professors at these institutions was 27.6 percent. I will speculate below that the greater share of non-tenure-track faculty members that is female is due to the difficulty that some female economists face in trying to combine tenure-track research careers and families at research universities.

Does the Falling Proportion of Tenured and Tenure-Track Faculty Matter?

A traditional argument for the importance of a tenure system for faculty is based upon academic freedom. Absent tenure, and the job security it provides, faculty members may be reluctant to pursue research on controversial issues. The importance of this rationale for tenure was brought home to me personally in the late 1970s

[6] Author's calculations from data in Tables 3 and 4 of the 2010 and Tables 2 and 5 of the 2005 CSWEP reports, available on the web at ⟨http://www.aeaweb.org/committees/cswep/annual_reports.php⟩.

when several trustees at my own institution challenged my promotion to professor because they disagreed with testimony I had given in a regulatory proceeding in the state of New York (as described in Ehrenberg 2002, p. 127). The Cornell Trustees shortly thereafter took the position, repeatedly affirmed, that the final decisions on tenure are to be made by the President and Provost of Cornell, with the Trustees only pro forma approving the decisions.

Economists have developed other arguments in support of tenure systems. One is that because a tenure system provides senior faculty with job security, they have an incentive to share their expertise with junior colleagues and students without creating competitors who will challenge their position; in this way, tenure facilitates the intergenerational transmission and expansion of knowledge (Stigler 1984). Another is that a tenure system can be thought of as an implicit long-term contract model, or a winner-take-all tournament model, and that both of these models can provide incentives for all faculty members to work harder (in the case of the contract model throughout the career; in the case of the tournament model, during the years prior to tenure and then to full professor) than would otherwise be the case (Lazear 1979; Rosen and Lazear 1981). In addition, a traditional labor economics argument holds that tenure is a desirable job characteristic and, in the absence of a tenure system, academic institutions would have to pay higher salaries to attract faculty. Indeed, in Ehrenberg, Pieper, and Willis (1999), my coauthors and I found that, ceteris paribus, economics departments that offer lower probabilities of tenure have to pay higher starting salaries to attract new faculty. A final argument is that if it is desirable for academics to specialize in their research in certain narrow subject areas, they need the reassurance of a reasonable probability of receiving tenure, because otherwise their specialization puts them at risk of having few alternative career options.

However, these arguments taken as a group seem to apply more to the role of faculty in research and institutional governance, rather than teaching. Is anything lost if undergraduate students are largely taught by adjuncts or full-time, non-tenure-track faculty, while a smaller number of tenure-track faculty focus on research and graduate education? After all, undergraduate students in most courses are typically being taught material that is far inside the research frontier. Does a more costly reliance on tenured and tenure-track faculty bring corresponding benefits for undergraduate education?

Only recently have economists and other social scientists begun to address this issue. While the results have been mixed, the existing research does suggest that a greater presence of tenured and tenure-track faculty will enhance undergraduate student outcomes. For example, in Ehrenberg and Zhang (2005), my coauthor and I used institutional-level panel data and found that—holding constant other variables including the socioeconomic backgrounds and test scores of entering students, and controlling for institutional fixed effects—when a four-year academic institution increases its use of either full-time, non-tenure-track faculty or part-time faculty, its undergraduate students' first-year persistence rates and graduation rates decrease. Using a similar methodology, Jacoby (2006) found that public two-year colleges that relied more heavily on part-time faculty had lower graduation rates, while Eagan and Jaeger (2009) and Jaeger and Eagan (2009) found that increased exposure of two-year college students to part-time faculty reduced the likelihood of the students transferring to four-year colleges or completing their associate's degrees. Finally, Bettinger and Long (2007) found that students attending Ohio public four-year colleges that take "adjunct heavy" first-year class schedules are less likely to persist in college after their first year; Jaeger and Eagan (2011) found a similar result for public two-year college students within a single state system.

In contrast, Bettinger and Long (2010) showed that having an adjunct as an instructor in an introductory class in some professional fields increases the likelihood that a student will take additional classes in the field, while Hoffman and Oreopoulos (2009) found that the tenure/tenure-track status and full-time/part-time status of a faculty member has no impact, on average, on student outcomes at a major Canadian research university. Of course, the costs of any increased use of non-tenure-track faculty on graduation and persistence rates must also be balanced against the financial savings from doing so. In Ehrenberg and Zhang (2005), we found, for example, that a 10 point increase in the percentage of full-time faculty not in tenure-track positions was associated with a 4.4 percentage point reduction in graduation rates at public master's-level institutions. As Table 3 indicates, the difference in average salaries between full-time lecturers and assistant professors at these institutions was over $10,000 a year in 2009–2010.

Given that many non-tenure-track faculty members are dedicated teachers and can devote themselves fully to undergraduate education because they face lesser research expectations, why might they be associated with lower student outcomes than their tenured and tenure-track faculty colleagues? One likely reason is that adjunct faculty appointments are often ad hoc in nature and instructors trying to eke out a living from this type of work must take on higher teaching loads, perhaps spread in across multiple institutions within an urban area, which leaves them little time and often no place to meet students outside of class.[7] Adjunct faculty in this difficult situation are also less likely to be up-to-date on their department's curriculum and may be less prepared to advise students. Non-tenure-track faculty who are full time will often have higher teaching loads than the teaching loads for the tenure-track faculty, which may also leave them with less time to work with individual students outside of class or to keep up with new developments in their field in a way that might encourage students to persist.

The increased pressure for faculty at major research universities to specialize in research has led the doctoral institutions to make greater use of full-time, non-tenure-track faculty in undergraduate education, especially at private universities (as shown earlier in Table 1). On the supply side, the relatively poor academic labor market conditions that currently confront new Ph.D.s, coupled with the large and growing salary differentials between major private research universities and

[7] Zhang and Liu (2010) show that four-year academic institutions in urban areas make more use of part-time faculty than other four-year institutions.

Table 3
Average Faculty Salary, by Rank and Institution Type in 2009–2010

Institution/Rank	*Assistant professor*	*Lecturer*	*Lecturer at private doc./ Asst. prof. in category*	*Lecturer at public doc./ Asst. prof. in category*
Private doctoral	83,573	61,860	0.74	0.62
Public doctoral	68,718	52,529	0.90	0.76
Private master's	63,003	55,272	0.98	0.83
Public master's	59,959	49,796	1.03	0.88
Private bachelor's	58,762	58,167	1.05	0.89
Public bachelor's	57,001	50,628	1.09	0.92
Two-year colleges	53,757	52,681	1.15	0.98

Source: American Association of University Professors, 2010, *The Annual Report on the Economic Status of the Profession: 2009–2010*, table 4. Available at: ⟨http://www.aaup.org/AAUP/pubsres/research/compensation.htm⟩.

virtually all other categories of academic institutions (Ehrenberg 2003), have made full-time, non-tenure-track teaching positions at the private doctoral universities an increasingly attractive alternative for new Ph.D.s.

This increased usage of full-time, non-tenure-track teaching positions has brought some efforts to improve the status of such faculty. While teaching loads of these faculty are often higher than those of their tenure-track colleagues (in part because the teaching loads of the latter have declined over time), teaching loads for the non-tenure-track faculty at the private doctoral universities are often lower—or at least no higher—than they would be if they were employed at other academic institutions in tenure-track positions.[8] For example, a fall 2003 survey found that while full-time instructional faculty and staff at public and private doctoral institutions spent an average of about 8 hours per week in the classroom, those at public and private master's programs spent about 11 hours per week in the classroom, and those at public two-year institutions spent 18 hours per week in the classroom (National Center for Education Statistics, 2005).

Table 3 presents data for 2009–2010 on average faculty salaries for assistant professors and lecturers (all departments), by institution type and form of control,

[8] The American Economic Association collects information on average yearly course loads for new assistant professors in economics departments in its annual *Universal Academic Questionnaire* and reports this information annually in its *American Economic Review* Papers and Proceedings (May) issue. The number of respondents to these questions is small, and the respondents vary over time. The responses (with sample sizes in parentheses) were, for academic year 1999–2000: Ph.D. institutions, 3.5 (30); master's institutions, 5.6 (11); and bachelor's institutions, 5.8 (38). For academic year 2009–2010, they were: Ph.D. institutions, 2.9 (35); master's institutions, 6.3(10); and bachelor's institutions 5.6 (32). The greatest reduction in course load was at the Ph.D. institutions. Charles Scott and John Siegfried provided me with information on ten departments at Ph.D. institutions that reported information for both 2010–11 and 2000–01, and the mean course load for new faculty at these departments fell during the ten-year period from 3.27 to 3.0 courses a year.

from a salary survey done by the American Association of University Professors. The private data are for non-church-related institutions. The average salary of lecturers at private doctoral universities is about $21,500 less than the average salary of assistant professors at those universities; however, it is only slightly lower than the average salary of assistant professors at public doctoral and private master's institutions and is higher than those of assistant professors at public master's, public and private bachelor's, and two-year colleges. These data suggest that the financial costs of accepting a lecturer position at a private doctoral university, if any, may not be that high relative to accepting an assistant professor position at most other types of academic institution, at least in the short-run. Furthermore, these non-tenure-track jobs need not come without a degree of job security. Conversations that I have had with economists at several private doctoral universities, who are either employed in non-tenure-track positions or are chairs of departments that hire such faculty, suggest that many of these positions now often offer "rolling multiyear contracts." For example, a lecturer may teach under a three-year contract that can be extended annually for a year if performance is satisfactory. Moreover, positions for non-tenure-track faculty members often have low or no research expectations, while offering an opportunity to teach at a major university with bright students and high-quality colleagues.

The data in Table 3 indicate that the average salary of lecturers at public doctoral universities is lower than the average salary of assistant professors in all categories of institutions. However, because of the lower teaching loads that the public doctoral institutions offer, jobs at such institutions may be attractive to new Ph.D.s given the current conditions of the academic labor market. Such programs also can attract high-quality, non-tenure-track faculty. Given the access of both public and private doctoral institutions to high-quality, non-tenure-track faculty, it should not be surprising that in Ehrenberg and Zhang (2005) we found that the expansion of full-time, non-tenure-track positions at doctoral universities had a smaller effect on undergraduate students' persistence and graduation rates than it had at the public master's-level institutions.

The data cited above from the Committee on the Status of Women in the Economics Profession (CSWEP) indicate that the average share of non-tenure-track faculty that is female at Ph.D. departments of economics is greater than the average share of assistant professors that is female at these same departments. A considerable body of research has noted the underrepresentation of females, relative to their share of new Ph.D.s, in tenure-track positions in science and engineering fields at research universities. A study by the National Research Council (2010) found that this underrepresentation is largely because female Ph.D.s are not applying for these positions at the same rate as their male counterparts. One obvious possible reason for this is that female scientists in their child-bearing years face a more difficult challenge than their male colleagues in striking a work–life balance (Mason and Goulden 2004). As a result, many research universities are adopting policies to alter the workplace and faculty culture to accommodate family issues (see for example, the UC Family Friendly Edge project, at ⟨http://ucfamilyedge.berkeley.edu⟩).

Table 4
Annual Average Percentage Real Changes in Expenditures per Full-Time Equivalent Student: FY1987–2008

	N	*Instruction*	*Student services*	*Academic support*	*Research*	*Public service*	*Institutional support*	*Operations and maintenance*	*Auxiliary*
Public doctoral	151	0.87	1.64	1.39	2.89	2.13	1.35	0.79	0.46
Private doctoral	103	1.87	3.13	2.87	2.35	2.83	2.60	1.05	1.42
Public master's	227	0.72	1.82	1.49	2.80	2.81	1.27	0.70	0.06
Private master's	327	1.55	2.66	2.13	2.18	0.75	1.57	–0.33	0.11
Private bachelor's	461	1.70	3.05	2.17	2.95	1.26	1.76	–0.23	0.52
Public 2-year	739	0.67	1.57	1.14	0.06	1.00	1.42	0.76	0.42
All public	1,192	0.75	1.66	1.22	2.74	1.69	1.39	0.77	0.37
All private	891	1.67	2.94	2.22	2.39	1.40	1.79	–0.12	0.49
All	2,083	1.07	2.16	1.62	2.63	1.66	1.57	0.51	0.40

Source: Author's calculations from Integrated Postsecondary Education Data System (IPEDS) data as cleaned by the Delta Cost Project (⟨http://www.deltacostproject.org⟩). Public bachelor's institutions are excluded from this table because of the relatively small number of them that reported data in both years.
Note: Institutions classified as "master's" typically offer undergraduate and master's degrees; and those classified as "doctoral" typically offer a wide range of undergraduate and graduate degrees including doctoral degrees.

Why Is a Declining Share of Resources Going to Instruction?

The share of academic resources going to instructional expenditures has declined at all categories of public and private not-for-profit institutions. On average, instructional expenditures per full-time-equivalent student—primarily faculty salary and benefits—increased by 1.07 percent a year above the rate of increase in the Consumer Price Index during the fiscal years 1987–2008, as shown in the bottom row of Table 4. In contrast, average real expenditures per full-time-equivalent student grew at more rapid annual rates for most other categories of institutional expenditures. These reallocations of funds away from instruction have been a major factor driving the shift away from full-time tenure and tenure-track faculty.

Why did these budget reallocations occur? The funding of higher education institutions comes from a variety of sources—and funds provided for some activities cannot be transferred to other activities.[9] For example, the "public service" category includes separately budgeted funds for non-instructional services to external

[9] Appendix Table A1, available online with this paper at ⟨http://e-jep.org⟩, illustrates how the sources of funds vary across categories of public and private not-for-profit academic institutions.

groups, such as cooperative extension activities, public broadcasting, and externally funded conferences. These activities are supported largely by targeted state appropriations, external grants, and targeted fundraising, and these funds cannot be used to support instructional activities.

Similarly, the research category includes sponsored research, which grew substantially during the period. Funds provided by external sponsors for research cannot be used for instruction. Moreover, during this period the share of academic research supported out of institutional funds grew dramatically, due to limitations established by the Office of Management and Budget in 1991 on "federal indirect cost rates" or "facilities and administration charges" (the "mark-up" allowed on the direct costs of research when universities are reimbursed through government research grants); growing requirements by the federal government for matching funds in grant proposals; and the growing cost of providing start-up funding for new scientists and engineers, which often is not recoverable in indirect cost rates (Ehrenberg, Rizzo, and Jakubson 2007). As a result, the percentage of academic institutions' total cost of research that is paid for by the institutions themselves out of institutional funds grew from about 12 percent in fiscal year 1976 to over 20 percent in fiscal year 2008 (Berdahl 2009). Increases in the institutional resources that academic institutions devote to research are associated, ceteris paribus, with increases in student/faculty ratios and with some substitution of full-time lecturers for professorial rank faculty (Ehrenberg, Rizzo, and Jakubson 2007). In addition, the growing use of part-time faculty at doctoral institutions, holding constant the use of full-time faculty, has been shown to be associated with increased external research and development expenditures at an institution; using adjuncts at doctoral universities to reduce the teaching loads of full-time faculty allows the full-time faculty to generate more external research funding (Zhang and Ehrenberg 2010).

"Student service expenditures" include costs of admissions, registrar activities, and activities whose primary purpose is to contribute to students' emotional and physical well-being and to their development outside of the classroom. Examples include student activities, cultural events, student newspapers, intramural athletics, student organizations, supplementary instruction (such as tutoring), and student records. Intercollegiate athletics and student health services *may* also be included in this category of expenses, except when they are operated as self-supporting auxiliary enterprises.[10] The annual growth rates of student service expenditures are roughly double those of the annual growth rates of instructional expenditures for every category of academic institutions.

These expenditures are viewed by some critics as discretionary "frills" that make no direct contribution to students' persistence in and graduation from college. In

[10] In the sample upon which Table 4 is based, over half of the institutions included varsity athletics in student service expenditures. The percentage that did so was over 80 percent for the private bachelor's and master's institutions and slightly more than 20 percent for the public doctoral institutions. "Public doctoral" is the only category of institutions in which more than a majority of the institutions (over 70 percent in our sample) included varsity athletics under auxiliary expenditures.

Webber and Ehrenberg (2010), we showed, however, that they do positively influence both first-year persistence rates and graduation rates of undergraduate students at four-year academic institutions. Moreover, as one might expect, these expenditures have greater effects at institutions that enroll a greater share of students who are disadvantaged, as measured by either their average entrance test scores or the levels of Pell Grant dollars that they receive. Indeed, our simulations suggest that at institutions whose graduation rates were below the mean in the sample, reallocating some resources from instruction to student services would lead, on average, to an increase in graduation rates; a similar reallocation was shown not to increase graduation rates at institutions whose graduation rates were initially at or above the mean. At least for a subset of higher education institutions, the more rapid growth of student service expenditures over the period may not be symptomatic of waste.

"Academic support expenditures" are for the activities and services that support instruction, research, and public service, including libraries, museums, and academic computing. The more rapid growth rate of expenditures in this category happens in part because, while the corporate world often adopts technology to cut costs, in the academic world, technology has often been adopted by academic institutions to enhance student learning and provide students with tools they will need to compete in the job market (Archibald and Feldman 2011). Another factor in this category is the growing costs of libraries; inflation rates for library materials have, for a long time, far exceeded the general rate of inflation, and the proliferation of electronic journals have increased, rather than decreased library costs (Ehrenberg 2002, chap. 14). The Association of Research Libraries (2009, table 2) reports that between 1986 and 2006, the average price of a serial purchased by research libraries increased by 5.3 percent a year; the average annual increase in the Consumer Price Index was 3.05 percent during the same period.

"Institutional support expenditures" include legal, finance, audit, human resources, budget, alumni affairs and development, audit and risk management, and public relations costs of the university. A dramatic proliferation of government regulations and reporting requirements, as well as a cap of 26 percent in the administrative cost component of federal indirect cost rates, has substantially increased the costs borne by academic institutions in this category. Higher education institutions regularly plead for regulatory relief and an easing of reporting requirements in a variety of areas, including human subjects, animal research, effort reporting, financial reporting, conflict of interest, and hazardous materials (Association of American Universities 2011).

Higher education institutions have increasingly devoted more resources to alumni affairs and development activities, seeking to enhance flows of giving from alumni, other individuals, corporations, and foundations. From fiscal years 1989 to 2009, voluntary support to higher education institutions per student grew, on average, by about 2.3 percent a year in real terms (Council for Aid to Education 2010, table 2). These funds support current operations, capital projects, and the endowment—so not all giving shows up in current operating budgets. While the costs of generating gifts varies widely, a widely cited 1990 study found that the mean cost over all academic institutions was in the range of 15 to 17 cents per dollar raised in the late 1980s (Council for Advancement and Support of Education, 1990). A new study is underway; the results from its pilot study of a relatively small number of institutions indicate that while the costs per dollar raised continue to vary across institutions, on average they remain similar to the earlier study.[11] If the costs per dollar raised remained roughly constant over the period, academic institutions' investments in fund-raising clearly also contributed to the increase in institutional support expenditures.

Expenditures on auxiliary enterprises are typically supported primarily by user fees: for example, hospitals, campus stores, residence halls, and food service all receive very little support from institutions' operating budgets. These expenditures, as well as those on operations and maintenance, grew at slower rates than instructional expenditures. Kaiser and Davis (1996) estimated that American higher education institutions had $26 billion dollars of accumulated deferred maintenance in 1995, of which $5.7 billion were urgent needs, so the slow growth of operations and maintenance expenditures may portend longer-run problems. Private conversations that I have had with James A. Kadamus, Vice President of Sightlines, a facility asset advisory firm that has the largest verified academic institution facilities database in the country, also suggests that this may well be the case. Academic institutions, in particular public institutions, have large aging facilities structures; recently, funding for maintenance of these facilities has not kept up with needs. And the additions of new facilities increases operating and maintenance needs, often without full thought in advance about where operating and maintenance funds will come from. Only a rare institution firmly commits not to increase the total square footage of facilities on the campus. However, the Ohio State University took this step in June 2010, when the Board of Trustees adopted a framework for capital facilities that called for adding to academic space only as replacements for existing facilities.

The explanations I have provided for the decreasing share of academic budgets going to instruction does not mean that I believe that academic institutions have always carefully controlled their administrative costs. They have not! Political scientist Benjamin Ginsberg (2011) argues that the growth of administration is largely due to the growth of a class of professional administrators who seek to "feather their own nests"; the result is the expansion of the bureaucracy and the declining role of the faculty in academic governance. However, the financial meltdown and deep recession that started in 2008 caused many colleges and universities to address their administrative cost levels. A number of the more wealthy public and private universities hired outside consultants to advise them how to restructure their administrative services (Keller 2010). The consultants' recommendations commonly fell into several main categories; reducing the layers of administration; increasing the number of direct reports each administrator supervises; centralizing procurement at large institutions and systems of institutions to achieve price concessions from

[11] Private correspondence from Rae Goldsmith, Vice President for Advancement Resources, Council for Advancement and Support of Education (CASE).

suppliers; achieving efficiencies in information technology; and reorganizing the delivery of support services, such as finance, communications, and human resources. At Cornell, for example, we expect to achieve savings of $75–85 million a year on our Ithaca campus by fiscal year 2015 from these efforts (see ⟨http://asp.dpb.cornell.edu⟩), which represents more than 5 percent of Cornell's operating budget once one removes external research funding. Continual efforts to reduce both administrative and other costs will be necessary if academia is to have any hope of reducing rates of tuition increases.

Changing Modes of Instruction, Technology, and the For-Profit Sector

The financial pressures being placed on academic institutions, along with demands to increase access and to support students in persisting to the completion of a degree, are forcing institutions to reexamine how they educate students. Institutions are reexamining the prevailing "lecture/discussion" format. Many institutions, in particular those in the for-profit sector, are seeking to use technology to improve learning outcomes *and* to reduce the cost of instruction, especially in remedial and introductory classes. Several evaluation studies suggest online education can be as effective as regular classroom contacts, especially for mature students, and that a blend of online and face-to face instruction is often more effective than online instruction alone (Means, Toyama, Murphy, Bakia, and Jones 2009). These efforts may well have substantial effects on costs and on the nature of the academic workforce in the future.

One prominent illustration of this point is the work of the National Center for Academic Transformation (NCAT), which has led efforts to use information technology to improve learning outcomes and reduce costs. The NCAT website ⟨http://www.thencat.org/⟩ lists over 30 large introductory courses that have been redesigned with its help, in quantitative, social science, humanities, and professional fields at a wide range of academic institutions, and provides links to descriptions of each redesign. The NCAT efforts tend to focus on replacing lectures with interactive computer-based learning resources such as tutorials, exercises, and frequent low-stakes quizzes, as well as individual and small group activities. Other points of emphasis include designing classes around mastering a series of learning objectives, and providing on-demand help, often in computer labs or online, staffed by a mixture of faculty, graduate assistants, peer tutors, or course assistants. Some of the cost reduction comes from a reduced reliance on costly full-time faculty and graduate assistants and an increased use of less-costly peer tutors and course assistants, who do things such as troubleshooting technical questions, monitoring student performance, and alerting the instructor to difficulties with teaching materials. This process may also enable institutions to leverage their best teachers more effectively.

Despite the successes of the classes reworked under the guidance of the National Center for Academic Transformation, dissemination of this model within and across institutions has been slow. There are numerous reasons: faculty skepticism about the usefulness of NCAT's approaches (in the face of the evidence); concerns about infringement on academic freedom in making decisions about how to teach; the unwillingness of some faculty to invest in new teaching methods; departmental concerns that the benefits of cost reduction will not accrue to them and that they will lose faculty positions; the difficulty of obtaining funds for required capital investments; and the need for stable leadership at departmental, college, university, and system levels committed to changing modes of instruction (Miller 2010).

A second example of innovative technology-based pedagogy comes from the Open Learning Initiative at Carnegie Mellon University (at ⟨http://oli.web.cmu.edu/openlearning/initiative⟩). This project has designed more than a dozen classes in introductory subjects—in primarily mathematics and science fields—that make use of advances in cognitive knowledge about how learning occurs and that use technology to create intelligent tutoring systems, virtual laboratory simulations, and frequent opportunities for assessment. The Open Learning Initiative has made these classes freely available on its webpage. An evaluation of an introductory statistics class taught at Carnegie Mellon showed that when a hybrid model that combined online learning with classroom instruction was used, students learned as much or more than they did in classes using traditional instructional methods and in half the time (Lovett, Meyer, and Thille 2008). Other evaluations have confirmed the effectiveness of the Open Learning Initiative approach for other classes and for students at large public universities and community colleges (Thille and Smith 2011).

The activities of both the National Center for Academic Transformation and the Open Learning Initiative suggest that technology can be used to improve educational outcomes and reduce the time (per student) spent by faculty in introductory-level classes at institutions ranging from community colleges to doctoral institutions. These initiatives appear less likely to influence methods of instruction in specialized upper-level elective classes. Their activities also suggest that the comparison that one should be making is not between lecture classes taught by adjuncts and those taught by tenured professors (as many of the studies I cited earlier implicitly did), but between the various different ways of organizing and staffing a course and the traditional lecture/discussion format.

More specifically, how has teaching of economics changed? National surveys have been conducted of the teaching methods used by academic economists in their classrooms in 1995, 2000, 2005, and 2010 (Watts and Becker 2008; Watts and Schauer 2011). While these surveys offer some evidence of increased use of PowerPoint displays, instructors putting class notes online, increased use of computer lab assignments in econometrics classes, and increased use of classroom experiments in introductory economics classes, the surveys also suggest that "chalk and talk" remains the dominant teaching method in economics (Watts and Schauer 2011).

However, response rates to the surveys have not been high: the 2010 survey had a response rate of only 10.5 percent. Thus, the surveys may not be capturing innovations in teaching economics. For example, there is an Open Learning Initiative introductory economics class developed by John Miller at Carnegie Mellon University that is associated with a textbook based in experimental economics

(Bergstrom and Miller 1999). An innovation from the private sector involves Aplia, an educational technology company founded in 2000 by Paul Romer, which offers online homework assignments (with immediate grading), math and graphing tutorials, articles from news sources, real-time online market experiments, and course management systems. Currently, Aplia offers course support for introductory and intermediate microeconomics and macroeconomics, as well as courses in money, banking, and financial institutions; international economics; and advanced placement economics. Many of these classes are integrated with leading textbooks in the field.[12] Lyssa Vanderbeck, Director of Program Management at Aplia, reported that about 147,000 students in over 4,900 economics courses used Aplia during the fall of 2010 (e-mail communication to me, June 1, 2011).

As demonstrated in Table 1, the growing for-profit higher education sector has been the leading sector in using part-time and full-time, non-tenure-track faculty. A growing number of institutions in this sector have also been in the forefront of attempting to use technology to improve educational outcomes and developing new methods of recruiting, training, and assessing faculty members. For example, the University of Phoenix, the largest for-profit, offers associate's, bachelor's, master's, and doctoral programs in primarily professional fields to primarily working adults. The vast majority of its faculty members are practicing professionals and part-time faculty. The University of Phoenix puts them through extensive orientation and training programs. About two-thirds of these faculty have master's degrees and one-third, doctoral degrees. Curricula are developed by experts and are fairly standardized. Extensive use is made of technology to facilitate student learning, including placing course materials online, using online tutors, and having students conduct their own online self-assessments of learning. Faculty members are evaluated both by feedback from students and from assessments of how well students have mastered the subject matter. As is common with most for-profits, University of Phoenix offers classes at times and places that are most convenient for students, especially working adults.

Institutions that compete most directly with the for-profits, in particular community colleges and comprehensive public universities, will increasingly face pressure to emulate the educational model of the for-profits: in particular, they will face pressure to expand their use of part-time faculty further and to consider evaluating faculty members based more upon student outcomes. At least so far, efforts by traditional academic institutions to embed student learning outcomes in course evaluations are few and far between. Examples of efforts to embed learning outcomes in course evaluations include those of the IDEA Center (at ⟨http://www.ideacenter.org⟩) and the Student Assessment of Learning Gains (at ⟨http://www.salgsite.org⟩).

[12] The Aplia website is at ⟨http://aplia.com/economics⟩. An example of Aplia's active learning materials are the active learning problem sets for principles of economics developed by Byron Brown of Michigan State University that are used by him in both his regular classroom and online teaching at ⟨http://www.bus.msu.edu/econ/brown/pim⟩.

Table 5
Percentages of Four-Year Public Institutions with Differential Undergraduate Tuition in 2010–2011

	Doctoral	*Master's*	*Bachelor's*
Number of institutions	172	271	120
Percent with any differential tuition	42	18	30
Percent with differential tuition by college or major	40	17	23
Percent with differential tuition by year enrolled	10	4	23

Source: Author's calculations from search of institutional web pages during the January to March 2011 period.
Note: Institutions classified as "associate's" typically offer two-year degrees as the highest degree; those classified as "bachelor's" offer primarily bachelor's degrees; those classified as "master's" typically offer undergraduate and master's degrees; and those classified as "doctoral" typically offer a wide range of undergraduate and graduate degrees including doctoral degrees. Bachelor's institutions exclude public colleges that offer some bachelor's degrees but that primarily offer associate's degrees.

Differential Tuition

American colleges and universities have historically charged the same tuition levels for all of their undergraduate majors (with the exception perhaps of laboratory fees). However, as Hoenack and Weiler (1975) and Siegfried and Round (1997) pointed out, an academic institution might plausibly seek to charge different tuition levels for different majors based upon the costs of providing an education in each major and the income-earning prospects that it offers. Indeed, a growing number of public institutions are adopting differential tuitions by college or major, or by year of enrollment. To gauge how prevalent this trend has become, from November 2010 to March 2011 my research assistants pored through the web pages of virtually all public academic institutions that grant bachelor's degrees searching for information on differential tuitions. Table 5 summarizes their findings. The percentage of public institutions with differential tuitions in 2010–2011 was highest, at 42 percent, at the doctoral institutions. If one further narrows the doctoral category to flagship doctoral institutions, the percentage increases to over half.

Differential tuition for these institutions is typically by college or by major, although a smaller percentage of them have differential tuition by year of enrollment, with upper-level students being charged more per credit hour than lower-level students. At the public master's institutions, differential tuition is almost always by college or major. In contrast, at the public bachelor's institutions, when differential tuition policies arise they are equally likely to be by college or major as by year of enrollment.

The most common programs for which differential tuition charges occur are business, engineering, and nursing. Examples of differential tuition charges in 2010–2011 that were obtained from institutional web pages include a $75 per engineering course fee at the University of Maine (a 9.4 percent increase over the in-state tuition of $801 for a three credit course), a $400 per credit hour additional tuition for business classes at Arizona State University (a 72 percent increase over the in-state per credit hour tuition of $557), and a $460 per semester nursing program fee at the University of Kentucky (a 10.7 percent increase over the in-state lower-division semester tuition of $4,305).

The possible consequences of differential tuition policies have not been empirically examined. Does differential tuition by major influence students' choice of major? Do higher tuition levels for upper-level students affect ultimate graduation rates? If such effects exist, are they larger for students from lower-income families? How might differential tuition charges interact with state and institutional financial aid policies?

Looking to the Future

Many faculty members will bemoan the decline of a golden age of American higher education, with its heavy reliance on tenured and tenure-track faculty. However, higher education is not immune to economic forces. The pressures that public and private colleges and universities face to expand enrollment, to increase graduation rates, and to limit future cost increases will likely only exacerbate the decline in full-time tenured and tenure-track faculty. Increasingly, academic leaders realize that how we teach our students must change, especially for remedial and introductory-level classes, and that technology must be employed to improve learning outcomes and reduce the per student costs of delivering instruction (Stripling 2011).

I am not noted for my ability to forecast the future, but I conclude with some personal speculations. The wealthy private and flagship public research universities and the leading private liberal arts colleges are in a world of their own. They will have access to the resources necessary to maintain full-time tenured and tenure-track faculty. They will increasingly employ technology in introductory-level classes in an effort to expand active learning and reduce costs, but in their case much of the cost savings will be directed to enhancing the quality of upper-division classes and furthering the research enterprise.

At research universities, the use of full-time, non-tenure-track faculty will likely continue and increase. For at least some new Ph.D.s, the combination of the pay levels at these institutions, their relatively low teaching loads (compared to other types of institutions), the low or nonexistent research demands, the possibility of rolling multiyear contracts, and the attractions of working at a large university will suffice to keep these non-tenure-track positions attractive. One result of this shift will be to free up more of the time of tenured and tenure-track faculty for research.

At the public and private regional doctoral universities, the public and private comprehensives, the other liberal arts colleges, and the two-year colleges, an ever-increasing share of faculty will not have doctoral degrees and will not be full-time on tenure-track lines. The use of technology and people in nonfaculty positions (like student assistants) to reduce costs and increase learning in remedial and introductory-level classes will likely occur much more rapidly at these institutions.

For all academic institutions, pressures for accountability surely will increase; academic institutions are increasingly being asked to provide information on assessing student learning outcomes as part of the accreditation process. Recent research by Arum and Roksa (2011) that concluded very little learning occurs in higher education for a large proportion of American students surely will add to these pressures. As such, one might expect to see an increased focus, especially in remedial and introductory classes, on evaluating faculty, at least partially, by their students' outcomes, as the for-profits do. This will put additional stresses on faculty/administration relations and faculty governance, especially at public campuses where collective bargaining contracts may specify faculty evaluation processes.

Few students who enter a Ph.D. program do so for the promise of financial rewards: other professional schools and alternate careers often promise higher annual earnings. Instead, students considering a Ph.D., especially those not considering degrees in science and engineering fields, have historically done so with the dreams of becoming a tenured faculty member and then pursuing a combination of research and teaching while participating in the governance of an academic institution. However, obtaining a Ph.D. has already become a less-attractive option in many fields, given the lengthening periods of time to complete the degree and the low levels of tenure-track hiring in the academic job market in recent years. Between 1979 and 2009, at U.S. universities, the share of new doctorates awarded to U.S citizens and permanent residents (among recipients with known nationalities) fell from 88 to 69 percent. By 2009, less than 40 percent of the new doctorates in economics were awarded to U.S. citizens or permanent residents (*2009 Survey of Earned Doctorates*, tables 16 and 19, available at ⟨http://www.nsf.gov/statistics/nsf11306/⟩).

The share of faculty positions that are not on the tenure-track, and perhaps not full-time either, along with the high fraction of such positions staffed by faculty without a doctorate, will likely further discourage American college students from going on for Ph.D. study. Moreover, as the share of full-time tenured and tenure-track faculty dwindle, this group will inevitably play a lesser role in the governance of the institutions of higher education.

■ *Financial support for the Cornell Higher Education Research Institute (CHERI) is provided by the Andrew W. Mellon Foundation and I am grateful to the Foundation for its support. Raj Kannappan, Andrew Key, Kristy Parkinson, Mirinda Martin, Douglas Webber, and Kenneth Whelan, all graduate and undergraduate research assistants at CHERI, assisted me in obtaining some of the data used in the paper, and I am also grateful for their help. Extraordinarily helpful comments on earlier drafts were provided by Charles Clotfelter, Gary Fethke, Daniel Hamermesh, Philip Lewis, and John Siegfried, as well as from the JEP editors. In places, this paper draws on Ehrenberg (2011).*

References

American Association of University Professors (AAUP). N.d. "Trends in Faculty Status, 1975–2007." AAUP Fact Sheet. http://www.aaup2.org/research/TrendsinFacultyStatus2007.pdf.

American Association of University Professors (AAUP). 2010. *The Annual Report on the Economic Status of the Profession: 2009–2010,* table 4. Available at: http://www.aaup.org/AAUP/pubsres/research/compensation.htm.

Archibald, Robert B., and David H. Feldman. 2011. *Why Does College Cost So Much?* New York: Oxford University Press.

Arum, Richard, and Josipa Roksa. 2011. *Academically Adrift: Limited Learning on American Campuses.* Chicago: University of Chicago Press.

Association of American Universities. 2011. "Regulatory and Financial Reform of Federal Research Policy: Recommendations to the NRC Committee on Research Universities." January 21, Available at: http://www.aau.edu/publications/reports.aspx?id=6900.

Association of Research Libraries. 2009. *ARL Statistics: 2007–2008.* Washington, DC: Association of Research Libraries.

Baum, Sandy, and Jennifer Ma. 2010. *Trends in College Pricing 2010.* New York: College Board.

Berdahl, Robert M. 2009. "Thoughts on the Current Status of American Research Universities: A Presentation to the National Academy's Board on Higher Education and Work Force." November 16. Available at: http://aau.edu/research/statement_speech.aspx?id=6898.

Bergstrom, Theodore, and James Miller. 1999. *Experiments with Economic Principles: Microeconomics.* New York: McGraw-Hill.

Bettinger, Eric P., and Bridget Terry Long. 2007. "The Increased Use of Adjunct Instructors at Public Universities: Are We Hurting Students?" In *What's Happening to Public Higher Education,* edited by Ronald G. Ehrenberg, 51–70. Baltimore: Johns Hopkins University Press.

Bettinger, Eric. P., and Bridget Terry Long. 2010. "Does Cheaper Mean Better: The Impact of Using Adjunct Instructors on Student Outcomes." *Review of Economics and Statistics* 92(3): 598–630.

Brewer, Dominic J., Eric R. Eide, and Ronald G. Ehrenberg. 1999. "Does It Pay to Attend an Elite Private College? Cross-Cohort Evidence of the Effects of College Type on Earnings." *Journal of Human Resources* 34(1): 104–23.

Council for Advancement and Support of Education. 1990. *Expenditures in Fund Raising, Alumni Relations, and Other Constituent (Public) Relations.* Washington, DC: Council for Advancement and Support of Education.

Council for Aid to Education. 2010. *2009 Voluntary Support of Higher Education.* New York: Council for Aid to Education.

Dale, Stacy, and Alan B. Krueger. 2002. "Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables." *Quarterly Journal of Economics* 117(4): 1491–1527.

Dale, Stacy, and Alan B. Krueger. 2011. "Estimating the Returns to College Selectivity over the Career Using Administrative Earnings Data." NBER Working Paper 17159.

Desrochers, Donna M., Colleen M. Lenihan, Jane V. Wellman. 2010. *Trends in College Spending: 1998–2008.* Washington, DC: Delta Cost Project. Available at: http://www.deltacostproject.org/analyses/index.asp.

Eagan, M. Kevin, and Audrey J. Jaeger. 2009. "Effects of Exposure to Part-Time Faculty on Community College Transfer." *Research in Higher Education* 50(2): 168–88.

Ehrenberg, Ronald G. 2002. *Tuition Rising: Why College Costs So Much.* Cambridge, MA: Harvard University Press.

Ehrenberg, Ronald G. 2003. "Studying Ourselves: The Academic Labor Market." *Journal of Labor Economics* 21(2): 267–87.

Ehrenberg, Ronald G. 2006. "The Perfect Storm and the Privatization of Public Higher Education." *Change* 38(1): 46–53.

Ehrenberg, Ronald G., ed. 2007. *What's Happening to Public Higher Education? The Shifting Financial Burden.* Baltimore: Johns Hopkins University Press.

Ehrenberg, Ronald G. 2010. "The Economics of Tuition and Fees in American Higher Education." In Vol. 2 of *The International Encyclopedia of Education* 3rd edition, edited by Barry McGraw, Penelope Peterson, and Eva Baker, 229–34. Oxford, UK: Elsevier.

Ehrenberg, Ronald G. 2011. "Rethinking the Professoriate." In *Reinventing Higher Education: The Promise of Innovation,* edited by Ben Wildavsky, Andrew Kelly, and Kevin Carey, 101–28. Cambridge, MA: Harvard Education Press.

Ehrenberg, Ronald G., Paul J. Pieper, and Rachel A. Willis. 1999. "Do Economics Departments with Lower Tenure Probabilities Pay Higher Faculty Salaries?" *Review of Economics and Statistics* 80(4): 503–12.

Ehrenberg, Ronald G., Michael J. Rizzo, and George H. Jakubson. 2007. "Who Bears the

Growing Cost of Science at Universities?" In *Science and the University,* edited by Paula E. Stephan and Ronald G. Ehrenberg, 19–35. Madison: University of Wisconsin Press.

Ehrenberg, Ronald G., and Liang Zhang. 2005. "Do Tenured and Tenure-Track Faculty Matter?" *Journal of Human Resources* 40(3): 647–59.

Ginsberg, Benjamin. 2011 *The Fall of the Faculty; The Rise of the All-Administrative University and Why It Matters.* New York: Oxford University.

Hoenack, Stephen A., and William C. Weiler. 1975. "Cost-Related Tuition Policies and University Enrollments." *Journal of Human Resources* 10(3): 332–60.

Hoffman, Florian, and Phillip Oreopoulos. 2009. "Professor Quality and Student Achievement." *Review of Economics and Statistics* 91(1): 83–92.

Jacoby, Daniel. 2006. "Effects of Part-Time Faculty Employment on Community College Graduation Rates." *Journal of Higher Education* 77(6): 1081–1103.

Jaeger, Audrey J., and M. Kevin Eagan. 2009. "Unintended Consequences: Examining the Effect of Part-Time Faculty Members on Associate's Degree Completion." *Community College Review* 36(1): 167–94.

Jaeger, Audrey J., and M. Kevin Eagan. 2011. "Examining Retention and Contingent Faculty Use in a State System of Higher Education." *Education Policy* 25(3): 507–37.

Kaiser, Henry H., and Jerry Davis. 1996. *A Foundation to Uphold: A Study of Facilities Conditions at U.S. Colleges and Universities.* Alexandria, VA: APPA, Association of Higher Education Facilities Officers.

Keller, Josh. 2010. "Universities Can Save Millions by Cutting Administrative Waste, Panelists Say." *Chronicle of Higher Education,* July 25. http://chronicle.com/article/Universities-Can-Save-Millions/123686/.

Lazear, Edward. 1979. "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87(6): 1261–84.

Lovett, Marsha, Oded Meyer, and Candace Thille. 2008. "The Open Learning Initiative: Measuring the Effectiveness of OLI Statistics Courses in Accelerating Student Learning." *Journal of Interactive Media in Education,* no. 14, JIME Special Issue: Researching Open Content in Education. http://jime.open.ac.uk/2008/14.

Mason, Mary Ann, and Mark Goulden. 2004. "Do Babies Matter (Part II)? Closing the Baby Gap." *Academe* 90(6): 11–16.

Means, Barbara, Yukie Toyama, Robert Murphy, Marianne Bakia, and Karla Jones. 2009. *Evaluation of Evidence-Based Practices in Online Learning: A Meta-Analysis and Review of Online Learning Studies.* Washington, DC: U.S. Department of Education.

Miller, Ben. 2010. "The Course of Innovation: Using Technology to Transform Higher Education." Education Sector Reports. May 18. http://www.educationsector.org/publications/course-innovation-using-technology-transform-higher-education.

National Association of College and University Business Officers. 2009. "Newly Released NACUBO Tuition Discounting Study Survey Report Shows Rates Remain Stable." http://www.nacubo.org/Research/Research_News/Newly_Released_NACUBO_Tuition_Discounting_Survey_Report_Shows_Rates_Remain_Stable.html.

National Association of College and University Business Officers. 2010. *2009 NACUBO Tuition Discounting Study of Independent Institutions.* Washington, DC: National Association of College and University Business Officers.

National Center for Education Statistics. 2005. Table 21 of *2004 National Study of Postsecondary Faculty (NSOPF:04): Background Characteristics, Work Activities, and Compensation of Instructional Faculty and Staff: Fall 2003.* December. http://nces.ed.gov/pubs2006/2006176.pdf.

National Research Council. 2010. *Gender Differences at Critical Transitions in the Careers of Science, Engineering and Mathematics Faculty.* Washington, DC: National Academy Press.

Rosen, Sherwin, and Edward Lazear. 1981. "Rank Order Tournaments as Optimal Labor Contracts." *Journal of Political Economy* 89(5): 841–64.

Scott, Charles, and John Siegfried. 2009. "American Economic Association Universal Academic Questionnaire Summary Statistics." *American Economic Review* 99(2): 641–45.

Siegfried, John, and David Round. 1997. "Differential Fees for Degree Courses in Australian Universities." In *Funding Higher Education: Performance and Diversity,* edited by Jonathan Pincus, and Paul Miller, 45–62. Canberra: Department of Employment, Education, Training, and Youth Affairs.

Snyder, Thomas D, and Sally A. Dillow. 2010. *Digest of Education Statistics 2009.* NCES 2010-13. Washington, DC: U.S. Department of Education.

State Higher Education Executive Officers. 2011. *State Higher Education Finance: FY 2010.* Boulder: SHEEO.

Stigler, George. 1984. "An Academic Episode." In *The Intellectual and the Marketplace.* Edited by George Stigler, 1–9. Cambridge, MA: Harvard University Press.

Stripling, Jack. 2011. "Governing Boards Turn to Technology to Reinvent the University." *Chronicle of Higher Education,* April 5.

Supiano, Beckie, and Andrea Fuller. 2011. "Elite Colleges Fail to Gain More Students on Pell Grants." *Chronicle of Higher Education,* March 27.

Thille, Candace, and Joel Smith. 2011 "Cold Rolled Steel and Knowledge: What Can Higher Education Learn About Productivity?" *Change* 43(2): 21–27.

Watts, Michael, and William E. Becker. 2008. "A Little More than Chalk and Talk: Results from a Third National Survey of Teaching Methods in Undergraduate Economics Courses." *Journal of Economic Education* 39(3): 273–86.

Watts, Michael, and Georg Schaeur. 2011. "Teaching and Assessment Methods in Undergraduate Economics: A Fourth National Quinquennial Survey." *Journal of Economic Education* 42(3): 294–309.

Webber, Douglas A., and Ronald G. Ehrenberg. 2010. "Do Expenditures Other than Instructional Expenditures Affect Graduation and Persistence Rates in American Higher Education?" *Economics of Education Review* 29(6): 947–58.

Zhang, Liang, and Ronald G. Ehrenberg. 2010. "Faculty Employment and R&D Expenditures at Research Universities." *Economics of Education Review* 29(3): 329–37.

Zhang, Liang, and Xiangmin Liu. 2010. "Faculty Employment at 4-Year Colleges and Universities." *Economics of Education Review* 29(4): 543–52.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2011—Pages 217–242

Compensation for State and Local Government Workers†

Maury Gittleman and Brooks Pierce

Are state and local government workers overcompensated? This question has taken on considerable heat in the last year or two, as many states and localities face budgetary stress. In this paper, we step back from the highly charged rhetoric and address this question with the two primary data sources for looking at compensation of state and local government workers: the Current Population Survey (CPS) conducted by the Bureau of the Census for the Bureau of Labor Statistics, and the Employer Costs for Employee Compensation (ECEC) microdata collected as part of the National Compensation Survey (NCS) of the Bureau of Labor Statistics. The fundamental difference between these two sources of data is that the CPS is a household survey while the ECEC is a survey of employers. Data from the NCS have been used in studies of union–nonunion, interindustry, and occupational wage differentials (Gittleman and Pierce 2007, 2011; Levenson and Zoghi 2011) and in studies of compensation inequality (Pierce 2001, 2010). However, while NCS publications regularly present tabulations separately for the private sector, on the one hand, and state and local government, on the other, the microdata from the NCS have not been previously used to compare compensation in the private sector to that in the state and local government sectors.

We begin by presenting some tabulations on pay differences across sectors from these two data series—both raw pay differences between public and private sectors and also some breakdowns by education level of workers (for the Current Population Survey) and skill level of the job (for the Employer Costs for Employee

■ *Maury Gittleman and Brooks Pierce are Research Economists, Bureau of Labor Statistics, Washington, D.C. Their e-mail addresses are ⟨gittleman.maury@bls.gov⟩ and ⟨pierce.brooks@bls.gov⟩.*

† To access the Appendix, visit http://dx.doi.org/10.1257/jep.26.1.217.
doi=10.1257/jep.26.1.217

Compensation series). We also discuss some important differences between these two data sets: in particular, the ECEC data has the considerable advantage that it includes costs of fringe benefits.

In both data sets, the workers being hired in the public sector have higher skill levels than those in the private sector, so the challenge is to compare across sectors in a way that adjusts suitably for this difference. We look at current wage and compensation gaps, trends in these gaps over time, and public–private differentials at different points of the distribution. We explore a number of methodological choices appearing in this literature, like whether or how to adjust for occupation or the size of an employer, and seek to understand which of these choices are important.

After controlling for skill differences and incorporating employer costs for benefits packages, we find that, on average, public sector workers in state government have compensation costs 3–10 percent greater than those for workers in the private sector, while in local government the gap is 10–19 percent. We caution that this finding is somewhat dependent on the chosen sample and specification, that averages can obscure broader differences in distributions, and that a host of worker and job attributes are not available to us in these data. Nonetheless, the data suggest that public sector workers, especially local government ones, on average, receive greater remuneration than observably similar private sector workers. Overturning this result would require, we think, strong arguments for particular model specifications, or different data.

Descriptive Statistics

Some tabulations for the raw wage gap between employees of state and local government and private sector employees are provided in Tables 1A and B. The first three columns of each table show the proportions of employment for state government, local government, and private sector workers, and then average hourly and weekly earnings by sector. In both datasets, the raw wage gap shows public sector workers being paid more. In the Current Population Survey data, the raw gap in hourly earnings is about 4 percent; in the Employer Costs for Employee Compensation data, hourly wages in the government sectors exceed those in the private sector by an average of about 30 percent.

To understand these numbers more deeply, it's useful to look more closely at the underlying data sources. The Current Population Survey is a monthly survey of about 60,000 households. In any given month, one adult household member reports employment and other information for each member of the household. A subset of households reports earnings and hours information. These are the "outgoing rotation groups," and each year since 1979 these interviews are gathered together into a single Merged Outgoing Rotation Group (MORG) file. The CPS-MORG includes demographic information on schooling and age, and information on jobs held such as industry, occupation, and the employer's sector, including state

Table 1A

Employment, Earnings, and Education by Ownership (Current Population Survey Data)

				Fraction with education level				
	Sample fraction	Average hourly earnings	Average weekly earnings	Less than high school	High school degree	Some college	College degree	Post-graduate
State government	.056	$22.55	$965.32	.017	.166	.246	.281	.290
Local government	.099	$22.33	$965.92	.026	.188	.256	.280	.251
Private sector	.845	$21.55	$960.58	.087	.300	.301	.217	.096
Combined sample	**1.0**	**$21.69**	**$961.41**	**.077**	**.281**	**.293**	**.227**	**.122**

Source: Based on data from the 2009 Current Population Survey, Merged Outgoing Rotation Group (CPS–MORG) file.
Notes: "Sample fraction" gives the hours-weighted sample proportion for each ownership group. Average weekly earnings are for full-time workers only. Sample size: 112,579.

Table 1B

Employment, Earnings, and Work Level by Ownership (National Compensation Survey Data)

				Fraction with work level				
	Sample fraction	Average hourly earnings	Average weekly earnings	Missing	1–4	5–8	9–12	13–15
State government	.039	$25.79	$1,016.70	.091	.197	.450	.241	.020
Local government	.101	$26.38	$1,055.07	.092	.281	.340	.284	.003
Private sector	.860	$20.18	$877.42	.107	.513	.268	.104	.007
Combined sample	**1.0**	**$21.18**	**$903.90**	**.105**	**.478**	**.283**	**.128**	**.007**

Source: Based on data from the Employer Costs for Employee Compensation (ECEC) in the 2009 National Compensation Survey (NCS).
Notes: NCS interviewers assign a level of work to all jobs in the survey, which ranges from 1 to 15, corresponding to pay levels in the General Schedule that sets levels of pay for federal workers. "Sample fraction" gives the hours-weighted sample proportion for each ownership group. Average weekly earnings are for full-time jobs only. Column entitled "Missing" provides proportion of jobs where work level information could not be obtained or in occupations that are considered "unlevelable." Sample size: 303,295.

government, local government, or private sector. For the CPS panel, Table 1A, the final five columns show a breakdown by education level. It's clear that state and local government employees are much more likely to have college degrees and postgraduate degrees than are private sector workers. Thus, later sections of this paper will seek to analyze what difference it makes to adjust these wage gaps by education level and other factors.

Table 1B shows descriptive statistics based on the Employer Costs for Employee Compensation, which is part of the National Compensation Survey (NCS). The NCS is a longitudinal establishment survey of nonfederal and nonagricultural employers. Interviewers visit newly sampled establishments and obtain information on the establishment and the jobs of a random sample of workers in the establishment. Jobs are defined using the employer's most narrow occupational classification or job title and other dimensions, including union coverage and full-time status. Information on individuals' earnings, job work schedules, and job work levels (described below and in the online Appendix available with this paper at ⟨http://e-jep.org⟩) is collected, but demographic information on job incumbents is not.

Because this survey captures information on the number of hours per week and weeks per year that employees in a job are scheduled to work, this information can then be used to convert earnings and compensation into hourly statistics. However, a potential problem arises here: the information on hours reflects employers' conceptions of scheduled work time. In most cases, work schedules are standard and easy to collect. However, an important exception for this study involves primary and secondary school teachers, whose actual hours worked per week are generally not available because time spent in lesson preparation, grading, and other nonclassroom activities are not available to employers. These data use the length of workday as specified by contract ("contract hours") for teachers in determining the work schedule, but given that, on average, teachers work more hours than their contract requires, the estimates of hourly earnings will be higher than if actual hours could be used (Schumann 2008; Allegretto, Corcoran, and Mishel 2004). Because of this measurement issue, we primarily analyze weekly earnings shown in the third column of the table, and we also restrict samples to full-time jobs (to help control for differences in weekly hours). When analyzing Employer Costs for Employee Compensation hourly earnings, we exclude certain occupations where contract hours are prevalent. We refer to this as "excluding teachers," even though that category includes much smaller job classifications in nonteaching occupations with collected contract hours, such as airline pilots, flight attendants, and others.[1] With the weekly earnings sample, state and local government earnings exceed those of the private sector by around 16–20 percent in the ECEC.

Although the National Compensation Survey does not collect information on education and experience by employee, the information on job levels can be used to compare pay at different skill levels across government and nongovernment jobs. This situation arises because the President's Pay Agent—an interagency group consisting of the Secretary of Labor and the Directors of the Office of Management and Budget and the Office of Personnel Management—uses the National Compensation Survey data to compare rates of pay for federal workers to nonfederal rates of pay, as called for by the Federal Employees Pay Comparability Act of 1990. For this reason, NCS interviewers assign a level of work to all jobs in the survey, which

[1] This does not exclude the entire educational sector. See the online Appendix available with this paper at ⟨http://e-jep.org⟩ for more information on measurement issues related to teachers' hours worked.

ranges from 1 to 15, corresponding to pay levels in the General Schedule system. In Gittleman and Pierce (2011), we demonstrate a close relationship between, on the one hand, education and experience in jobs in the Current Population Survey, and on the other, the factors underlying the work levels assigned to jobs by the National Compensation Survey interviewers.

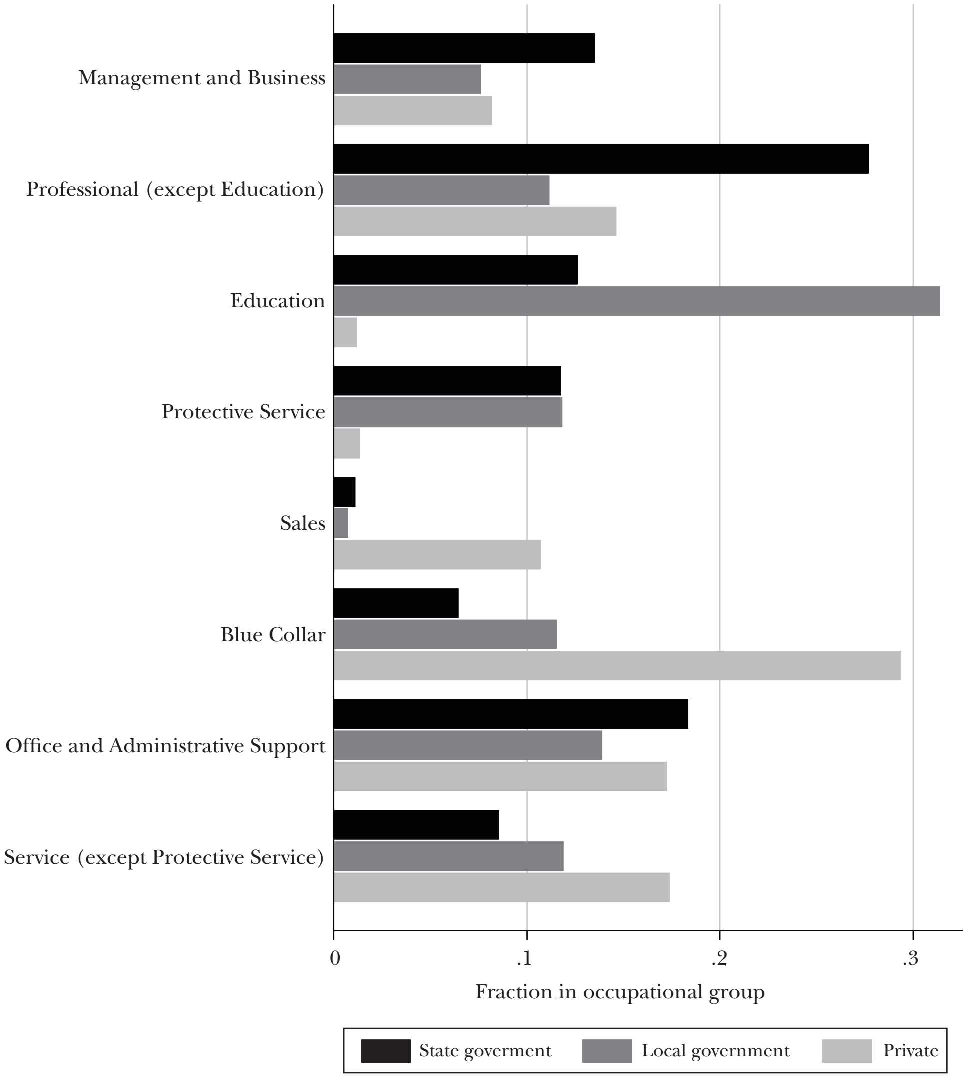
In the last few columns of the Table 1B, we first divide work levels into four categories, though only a small amount of employment is at the highest work levels (13–15). It is immediately apparent that a far higher fraction (0.513) of employment in the private sector is in the bottom category of work levels (1–4) than is the case for the local (0.281) and state governments (0.197). Again, this means that controlling for skill differences across sectors should substantially reduce estimated pay gaps.

Figure 1 presents, for the ECEC, the distribution of employment across eight occupational groups, formed by aggregating the 22 two-digit occupations in the Standard Occupational Classifications.[2] To focus on occupations that are important in government, Education remains split out from other Professional occupations, while Protective Service is kept separate from other Service. The occupational groups are ordered from high earnings at the top to low earnings at the bottom, based on earnings from all sectors combined. Consistent with the skill differences evident in Table 1A and B, the government proportions tend to be higher in the occupations at the top, and the private shares greater in those occupations at the bottom. Highlighting the difficulty of making private–government comparisons, even at this level of aggregation, some occupations, such as sales, are virtually all private, while others are almost all public. Differences between state and local government are also evident, with state government having higher concentrations in Professional and Management, and local government being disproportionately represented in Education. As one might expect, this coarse level of aggregation hides some interesting distinctions. For example, within the Education group, employment is relatively concentrated in kindergarten and preschool for the private sector, in primary and secondary teaching for local government, and in postsecondary teaching for state government.

Along with data on wages, the Bureau of Labor Statistics also collects the information on benefit costs necessary to compile the Employer Costs for Employee Compensation data for roughly half of the National Compensation Survey sample.

[2] The 22 two-digit Standard Occupation Classification codes are mapped into eight groups in Figure 1 as follows. *Management and Business* includes 1) Management and 2) Business and Financial Operations. *Professional, except Education* includes 1) Computer and Mathematical; 2) Architecture and Engineering; 3) Life, Physical, and Social Science; 4) Community and Social Services; 5) Legal; 6) Arts, Design, Entertainment, Sports, and Media; and 7) Healthcare Practitioner, and Technical. *Education* has no official subcategories but includes jobs in Education, Training, and Library. *Protective Service* has no subcategories. *Sales* has no subcategories. *Blue Collar* includes 1) Farming, Fishing, and Forestry; 2) Construction and Extraction; 3) Installation, Maintenance, and Repair; 4) Production; and 5) Transportation and Material Moving. *Office and Administrative Support* has no subcategories. *Service, except Protective Service* includes 1) Healthcare Support; 2) Food Preparation and Serving Related; 3) Building and Grounds Cleaning and Maintenance; and 4) Personal Care and Service.

Figure 1
Occupation Distributions by Sector



Source: Employer Costs for Employee Compensation (ECEC) in the 2009 National Compensation Survey (NCS).

Broadly speaking, this includes employer costs associated with paid leave, health and other insurance plans, retirement and savings plans, certain forms of supplemental pay, and legally required benefit costs (such as Social Security), but excludes costs associated with retiree health plans. The ECEC data are converted to a cost-per-hour-worked average for incumbents within a job. It should be noted that these data can be difficult for a respondent to report. As an example, there is a certain amount of measurement error involved in getting job-specific data for some of the components of the ECEC because respondents are sometimes able to report data only for a broader group than the job incumbents, such as the average for all white-collar workers or for all workers. To give another example, it is quite difficult to price out the defined benefit pension obligations associated with current employees, and the NCS typically reverts to employer expenditures, which will depend on account funding rules and plan asset investment returns. It is possible that these kinds of measurement errors are not randomly distributed across the data, but, instead, that certain errors are more common in certain sectors.[3]

It is commonly understood that nonwage benefits form an important part of public sector compensation packages. Government workers are much more likely to be offered health insurance and retirement plans, and are more likely to enroll in such plans if offered. In addition, public sector plan structures tend to offer more comprehensive coverage. Public sector health plans tend to require lower employee contributions and have higher employer premiums, and are more likely to come bundled with supplemental dental, vision, or prescription drug plan components. Private sector retirement plans, when offered, are typically defined contribution plans rather than higher-cost defined benefit plans. Furthermore, differences exist within retirement plan type; for example, public sector defined benefit plans are more likely to include post-retirement cost-of-living adjustments. Public workers are also more likely to be eligible for retiree health benefits.[4]

Table 2 shows that such qualitative differences factor into employer costs. As the table indicates, the costs per hour worked for the various benefits collected are much greater in the public sector (about $14) than in the private sector (around $8). Spending on health insurance in the government ($4.30 at the state level and $4.56 at the local level) is more than double that in the private sector ($2.14), while expenditures on retirement and savings are more than triple ($3.18 and $3.37 versus $1.00). Within retirement and savings, the vast majority of spending in government goes toward defined benefit plans, while in the private sector, the breakdown is much more even between defined benefit and defined contribution plans. Paid leave is also more generous in government, more than double the private sector level in state government and more than 50 percent higher in local government. For the remaining category of "other benefits," which includes nonproduction bonuses, short- and long-term disability benefits, and all legally required benefits, the private sector has an edge. Two explanations for this are that bonuses tend to be higher in the private sector and that not all government workers are covered by Social Security.

[3] Another proviso is that employer costs for wages and benefits will differ from employee valuations of those same wages and benefits due to a number of considerations: taxes; the fact that benefits are not always easily adjustable to a given worker's desired level; and any divergence between an employer's price for a benefit and what an employee would have to pay as an individual (Famulari and Manser 1989).

[4] See ⟨http://www.bls.gov/ncs/ebs/benefits/2010/ownership.htm⟩ (Bureau of Labor Statistics) for supporting statistics.

Table 2
Employer Costs per Hour Worked and as a Percentage of Compensation, by Sector

	Average cost per hour worked ($/hr)			*Percent of compensation (%)*		
Compensation component	*Private*	*State govt.*	*Local govt.*	*Private*	*State govt.*	*Local govt.*
Wages and salaries	20.37	25.79	26.38	71.6	64.5	66.1
Total benefits	8.08	14.18	13.50	28.4	35.5	33.9
Health insurance	2.14	4.30	4.56	7.5	10.8	11.4
Retirement and savings	1.00	3.18	3.37	3.5	8.0	8.5
Defined benefit	0.43	2.65	3.09	1.5	6.6	7.7
Defined contribution	0.57	0.54	0.28	2.0	1.3	0.7
Paid leave	1.96	3.98	2.95	6.9	10.0	7.4
Vacation	1.02	1.82	1.10	3.6	4.6	2.7
Holiday	0.63	1.24	0.88	2.2	3.1	2.2
Sick	0.24	0.77	0.75	0.8	1.9	1.9
Personal	0.07	0.15	0.22	0.3	0.4	0.6
Other benefits	2.94	2.64	2.57	10.3	6.6	6.5
Excluding teachers						
Wages and salaries	20.18	23.42	22.69	71.6	63.3	63.6
Total benefits	8.02	13.60	12.97	28.4	36.7	36.4

Source: Authors' estimates based on employer costs from the 2009 Employer Costs for Employee Compensation sample of the National Compensation Survey.
Notes: Estimates are hours-weighted statistics. "Wages and salaries" include wages, commissions, piece rates, overtime pay, and shift differentials, but do not include nonproduction bonuses. "Other benefits" include nonproduction bonuses, short- and long-term disability benefits, and all legally required benefits. "Excluding teachers," refers to the exclusion of certain occupations where contract hours are prevalent; this excludes much smaller job classifications in nonteaching occupations with collected contract hours, such as airline pilots and flight attendants, but does not exclude the entire educational sector. See the online Appendix for details.

We caution that different benefits profiles do not automatically indicate that public sector plans are "too generous." It seems plausible that public sector workers demand a compensation package skewed more towards benefits. After all, they have higher incomes and are older and more educated. More speculatively, they may have preferences such as greater risk aversion or different rates of time preference that would induce greater demand for retirement and health insurance benefits. There may be contract design issues related to optimal retirement dates and specific human capital accumulation that make defined benefit plans more sensible for public sector workers. Furthermore, public sector employers almost certainly find plan provision cheaper due to economies of scale—for example, by avoiding some adverse selection issues in health insurance that plague small private firms.

However, one clearly must consider benefits together with wages; as such, we believe that the Employer Costs for Employee Compensation data (from the National Compensation Survey) contain valuable information for the problem at hand. They come from a representative sample, comprehensively cover the benefit spectrum, and are derived from employer and administrative records. While more work is undoubtedly needed on reconciling household and establishment surveys, we believe the National Compensation Survey likely contains more accurate data on wages, industry, occupation, and sector than does the Current Population Survey, which seems more subject to a number of possible error sources such as respondent error and imputations due to partial nonresponse.[5] One would expect more accurate information to lead to sharper estimated sector differentials.[6]

Making Comparisons: Approaches

When comparing the pay of two different groups, the economics literature has typically followed one of two paths (Moulton 1990; Belman and Heywood 2004b). One approach is the "opportunity wage definition" of comparability, sometimes also called the "people approach," in which workers with given characteristics in the government sector should be paid the same as they would be in the private sector. The alternative is a "positions approach," in which one would search the private sector for positions that match the descriptions of those in the public sector and then compare compensation, ignoring the characteristics of those who actually hold the positions. As Moulton (1990) has observed in the case of federal/nonfederal comparisons, the two approaches need not provide an answer of the same sign, let alone similar magnitudes.

Our approach depends to some extent on the dataset we are employing. With the Current Population Survey, we adopt an opportunity wage or a "people approach," because this household survey source lets us use demographic data. However, we can also add a control variable for occupations, which makes this approach something of a hybrid. The Employer Costs for Employee Compensation dataset, however, is a sample of jobs rather than of individuals. Thus, when we are using these data, we are primarily examining the characteristics of positions.

The two main regression-based approaches in the literature are the Oaxaca–Blinder decomposition[7] and a dummy variable approach, where wages are a function of the covariates and indicator variables for the different sectors. The two

[5] For example, an implausible 10.1 percent of Current Population Survey state government workers are in elementary, middle, and secondary school teaching occupations, versus 0.3 percent in the National Compensation Survey, suggesting some difficulty in CPS reports or imputations of sector of work or occupation.

[6] In Gittleman and Pierce (2011), we find sharper industry and occupation differentials in the National Compensation Survey. In addition, in that paper, we report R^2 values approaching 0.8 from wage regressions on the factors underlying work level, which is much higher than analogous wage regressions using the Current Population Survey.

[7] Fortin, Lemieux, and Firpo (2010) give an excellent exposition of this and other decomposition methods. There are, of course, more sophisticated approaches than either the Oaxaca–Blinder decomposition or the dummy variable approach that one could use to take account of unobserved heterogeneity and selectivity into the sectors (examples are Gyourko and Tracy 1988; Krueger 1988a; Lee 2004). The logic of such models does not carry over as well to a job-based dataset such as the Employer Costs for Employee Compensation. Moreover, these approaches rely heavily on the appropriateness of certain identifying assumptions (Moulton 1990; Gregory and Borland 1999). We do not attempt them here.

give similar results, and so, in this section, we focus on the first approach. For the purposes of the Oaxaca–Blinder decomposition, we specify that

$$\ln(y_{jk}) = \mathbf{X}_{jk}\boldsymbol{\beta}_j + \varepsilon_{jk},$$

where for sector j, which can be state, local, or private, and individual k, y_{jk} is either a wage or compensation measure, $\mathbf{X}_{jk}$ is a vector of characteristics and $\boldsymbol{\beta}_j$ is the corresponding coefficient vector for the relevant sector, and ε_{jk} is an error term with mean zero and variance σ_j^2. We then use the coefficients from this regression to estimate the log wage or log compensation that the average public sector worker would earn in the private sector, either state or local. Denoting the government and private sectors with subscripts g and p respectively, differences in sector-average log wages or compensation would be decomposed into an explained portion or composition effect $(\overline{\mathbf{X}}_p\hat{\boldsymbol{\beta}}_p - \overline{\mathbf{X}}_g\hat{\boldsymbol{\beta}}_p)$, and an unexplained remainder $(\overline{\mathbf{X}}_g\hat{\boldsymbol{\beta}}_p - \overline{\mathbf{X}}_g\hat{\boldsymbol{\beta}}_g)$, the wage structure effect. This unexplained portion shows how much more or how much less an average public sector worker would earn in the private sphere (absent any general equilibrium effects on the private wage structure associated with workers changing sectors). We therefore take what a public worker would earn in the private sector to be the relevant counterfactual to the worker's actual public sector earnings. When discussing the main results, we provide the differential in log points, though we also report differentials transformed into percentage terms in the tables.

What control variables are appropriate in this regression? Linneman and Wachter (1990), Hirsch, Wachter, and Gillula (1999), and others have argued that it is important to distinguish between skill-related factors that an individual can transfer from job to job and a second set of variables that are descriptive of the job or sector and possibly indicative of noncompetitive pay differentials such as rent-sharing. It is, of course, not always clear whether a variable falls in one camp or another. In regressions using the Current Population Survey data, we control for schooling and work experience in a more-or-less standard set of covariates. In regressions using data from the Employer Costs for Employee Compensation, we control for differences in human capital via a series of dummy variables for work level. Therefore, we assume that individuals, on average, possess the requisite skills to fill the positions and that these skills would carry over to the private sector.

We treat union status and organizational size as not reflecting worker skills. Controlling for union coverage seems inappropriate, because union wage premia probably do not reflect ability differences, and those in the public workforce would not likely take their public sector unionization rates with them if they were to move to the private sector. The situation is murkier in the case of employer size, because less of a consensus exists as to causes of the size premium. The traditional explanation has been that larger employers have greater product market power, and that workers capture some of these rents. If, however, employee compensation rises with the size of employer because larger employers hire better-quality workers—that is, employer size is a proxy for unobserved worker ability, even in the public sector—then including size as a control is desirable. Troske (1999) tests several explanations of the employer size-wage effect and a significant unexplained premium remains. This and other evidence leaves the door open for the possibility that rent-sharing may be involved. Absent evidence that larger public sector organizational size reflects unobserved ability, we do not control for employer size.[8] Estimates of differentials are sensitive to these choices and we will provide measures of this sensitivity.

Another issue on which the literature on public sector differentials has not reached a consensus is whether to account for occupation and, if so, at what level of detail. Of recent studies of state and local government differentials, Schmitt (2010) and Keefe (2010) use no occupational controls, while Bender and Heywood (2010) use them only in robustness checks. On the other hand, in Moulton's (1990) study of federal wage differentials, he argues that it is essential to control for occupation at as detailed a level as possible. Over the years, the modal choice has probably been one-digit or major occupation controls (Belman and Heywood 2004b). If occupation reflects unmeasured human capital and working conditions common to the private and public sectors, then including occupation controls will help net out cross-sectoral differences in these wage-influencing factors. But as Belman and Heywood discuss, even if one believes that occupation controls are important, it is not obvious how fine one should go: coarse controls may leave occupations too heterogeneous, while finer controls can remove unique occupations from the analysis. Given the lack of consensus, we present models with different treatments for occupation.

What is the Current Private–Public Pay Gap?

Estimates Based on Current Population Survey Data

We display the results of a comparison of private and public sector workers in Table 3, using weekly earnings for full-time workers in the Current Population Survey in the first pair of columns, and weekly earnings and compensation for full-time jobs in the Employer Costs for Employee Compensation from the National Compensation Survey in the second and third pairs of columns. The first column in each pair in Table 3 presents the log differences and, the second, the percentage differences, where the percentage differences are relative to public sector statistics. In this section, we discuss differentials in terms of log points because the Oaxaca–Blinder decomposition is in terms of logs. In later sections, we adopt a somewhat looser approach and only display percentage differences.[9] Asterisks are used to indicate statistical significance at the 1 and 5 percent levels.

[8] Brown and Medoff (1988) conclude that measured ability to pay accounts for about 15 percent of the public sector size–wage effect.

[9] We present both log differentials and percentage differences because they measure different things. Researchers often approximate percentage differentials by exponentiating log differentials. But, as stressed in Blackburn (2008), this standard transformation gives misleading results if wage dispersion differs substantially across sectors. It is well known that earnings distributions are less dispersed in the public than in the private sector (for example, Poterba and Rueben 1994), which would cause

Table 3
Private–Public Pay Differentials for Weekly Wage and Compensation

	CPS weekly wage		*NCS weekly wage*		*NCS weekly compensation*	
	logs	*percent*	*logs*	*percent*	*logs*	*percent*
A. Private–State Government						
Raw differential	–0.085**	–0.5	–.203**	–13.7**	–.335**	–21.8**
Unexplained differential						
Base controls, including education (CPS) or work level (NCS)	0.108**	16.2**	.048**	8.1**	–.076**	–3.2*
Plus major occupation	–0.005	2.7**	–.012	1.3	–.143**	–10.1**
Plus detailed occupation	0.020*	4.9**	–.003	2.3	–.125**	–8.7**
B. Private–Local Government						
Raw differential	–0.097**	–0.6	–.246**	–16.8**	–.347**	–23.0**
Unexplained differential						
Base controls, including education (CPS) or work level (NCS)	0.080**	13.6**	–.054**	–2.5**	–.149**	–10.5**
Plus major occupation	–0.091**	–5.2**	–.123**	–9.6**	–.236**	–18.5**
Plus detailed occupation	–0.067**	–3.5**	–.115**	–9.2**	–.220**	–17.6**

Sources: Based on data from the 2009 Current Population Survey, Merged Outgoing Rotation Group (CPS–MORG) file and data from the Employer Costs for Employee Compensation (ECEC) sample in the 2009 National Compensation Survey (NCS).
Notes: Samples are restricted to full-time workers or full-time jobs for these models. Estimates are private–public differentials in log points and percentage differences. Unexplained differentials are from Oaxaca-Blinder decompositions that net out sectoral differences in controls, using private sector returns. Base controls in the CPS include sex, Census division interacted with metropolitan area, four education dummies, and a quartic in experience. Base controls in the NCS include dummy variables for work level, and Census division interacted with metropolitan area. When detailed occupation controls are included, the raw differentials sometimes deviate slightly from the raw differential provided because occupations for which there are government workers but no private sector workers are excluded.
** and * indicate statistical significance at 1 and 5 percent levels, respectively.

Beginning with estimates based on the Current Population Survey Data, we see that the raw average earnings of state workers exceed those of private workers by about 0.09 log points and that the earnings of local government workers surpass those of private workers by about 0.10 log points. However, after including the baseline controls for education, experience, sex, full-time status, and the interaction of Census division with metropolitan area,[10] it is *private sector workers* who appear to

the standard transformation to understate private sector relative to public sector pay. See the online Appendix available with this article at ⟨http://e-jep.org⟩ for details on how we use log wage regression results to derive estimated differentials in percentage terms.

[10] Interacting Census division with metropolitan area is more flexible than just including Census division and metropolitan area dummies by themselves in that the metropolitan–nonmetropolitan differential is allowed to vary by division.

earn more than their state and local government equivalents, by 0.11 and 0.08 log points respectively. The implication of this regression is that public sector workers earn more on average than private sector workers because public sector workers have higher levels of human capital, particularly education—not because they receive higher pay for a given level of human capital.

But if one takes into account major occupation differences using the two-digit level of the Standard Occupational Classification system (which contains 22 occupations), the estimates once again change markedly. With a swing of 0.171 log points from the base control case, relative pay estimates now favor local government workers, who are paid 0.091 log points more than can be explained by the controls. For state workers, the shift from the base control case is smaller (0.113 points), so that the gap in log points virtually disappears. Our results using major occupational controls indicate that, conditional on the baseline variables—especially schooling levels—government workers are more likely to be in lower-paying two-digit occupations than their private sector counterparts. College-educated government workers are in less-lucrative two-digit occupations (like teaching) than their private counterparts, who are more likely to be healthcare practitioners or in areas like business and management. On the other end of the educational spectrum, those without a high school degree in the private sector are more likely to be in the relatively lucrative (conditional on schooling) production and construction occupations.

Moving from two-digit occupational controls to the most detailed occupational controls contained in the Current Population Survey—consisting of nearly 500 occupations—relative pay differentials shift about 0.02–0.03 log points in favor of private sector workers, so that private sector workers are 0.020 log points above the state government workers, while local government workers are about 0.067 points above the private sector. Of course, the models with and without occupational controls represent very different thought experiments. Readers who believe it likely that, say, college-educated teachers and managers have different levels of unmeasured human capital will tend to gravitate toward the models controlling for occupation. They will prefer to compare, conditional on schooling and other factors, teachers to teachers, and indeed they may prefer to compare elementary school teachers to elementary school teachers via detailed occupational controls. The alternative view is that it is more useful to make across-sector comparisons unconditional on occupation: perhaps occupation does not accurately reflect unmeasured skills, and occupational controls only exacerbate difficulties from (say) unmeasured differences in pecuniary or nonpecuniary factors.

All in all, the data from the Current Population Survey does not provide an unambiguous answer to the question of whether comparable workers receive higher wages in the public sector. Results differ by specification, and local government workers appear to generally fare better than state government workers. More importantly, however, the Current Population Survey does not contain comprehensive information on nonwage benefits. For that, we need to turn to the Employer Costs for Employee Compensation microdata.

Estimates Based on the Employer Costs for Employee Compensation Data

The baseline controls used in the regressions with the NCS Employer Costs for Employee Compensation microdata are as similar as possible to those in the Current Population Survey. However, in these regressions, instead of controlling for education and experience, we control for job work levels.

As already noted in Table 1, the baseline raw differentials are wider in the Employer Costs for Employee Compensation (ECEC) microdata. As shown in the third column of Table 3, the state and local government raw weekly earnings gap is 0.203 and 0.246 points, respectively, above the private sector. Using the base controls, which in this case include job work levels, private sector jobs with the same characteristics as state sector jobs pay about 0.048 log points more, which is smaller than the 0.108-point difference based on Current Population Survey data in column 1. For local government jobs, the ECEC results suggest an edge in pay for public sector workers of 0.054 points, versus the estimate of a 0.080-point advantage for the private sector in column 1. In general, the explained portions of the raw wage gaps are larger in the ECEC than in the CPS, which suggests that the work-level information provides more information about skill differences across jobs than is apparent in the demographic data about workers in the Current Population Survey. However, the remaining unexplained wage gap between private sector and local government workers indicates that local government workers are paid higher wages, and this is a departure from CPS-based studies that routinely find the opposite.

Because the work-level data were designed to make different jobs comparable, adding occupational controls does not affect the measured public sector premia as much in the "NCS weekly wage" column as it did with the Current Population Survey data in the first set of columns, where education level is the main control. With detailed occupational controls, there is little ground between the state and the private sector, but a difference of 0.115 log points remains in favor of the local government.

What effect do the more generous benefits provided to government workers have on private–public differentials? When nonwage compensation is included, as in the "NCS weekly compensation" columns of Table 3, the raw differentials widen markedly in favor of the state and local government workers. The state differential widens by 0.132, to 0.335 points, and the local differential by 0.101, to 0.347 points. When compensation is the dependent variable, the regression-adjusted estimates move (relative to the NCS weekly wage column) by almost as much as the raw differentials in favor of government workers. This shift tends to fall in the range of 0.10–0.12 log points.

Are government jobs more highly compensated than corresponding private sector jobs? The answer from the Employer Costs for Employee Compensation microdata appears to be "yes," although the magnitude of the difference depends upon sector of government and specification used.[11]

[11] Nonwage compensation in state and local government exceeds that of the private sector. Is there a way to determine which particular benefit categories contribute to this difference, conditional on job characteristics? The decomposition methodology in logs used here is ill-suited for this task since some

Table 4
Private–Public Percentage Pay Differentials for Hourly Wage and Compensation

	CPS hourly wage (%)	*NCS hourly wage (no teachers) (%)*	*NCS hourly compensation (no teachers) (%)*
A. Private–State Government			
Raw differential	–4.4**	–13.8**	–23.8**
Unexplained differential			
Base controls, including education (CPS) or work level (NCS)	12.4**	9.1**	–3.7*
Base controls and major occupation	0.7	3.5*	–9.2**
Base controls and detailed occupation	2.4*	2.5	–9.5**
B. Private–Local Government			
Raw differential	–3.5**	–11.1**	–20.9**
Unexplained differential			
Base controls, including education (CPS) or work level (NCS)	11.0**	–2.3**	–12.9**
Base controls and major occupation	–5.7**	–7.2**	–18.2**
Base controls and detailed occupation	–4.8**	–4.8**	–14.8**

Sources: Based on data from the 2009 Current Population Survey, Merged Outgoing Rotation Group (CPS–MORG) file and data from the Employer Costs for Employee Compensation (ECEC) sample in the 2009 National Compensation Survey (NCS).
Notes: Estimates are private–public differentials in percentage differences. Unexplained differentials are from Oaxaca–Blinder decompositions that net out sectoral differences in controls, using private sector returns. Base controls are as in Table 3, but also include indicators for full-time workers or jobs. When detailed occupation controls are included, the raw differentials sometimes deviate slightly from the raw differential provided because occupations for which there are government workers but no private sector workers are excluded.
** and * indicate statistical significance at the 1 and 5 percent levels respectively.

Hourly Sample Results and Comparisons with Other Studies

In Table 4, we present the results of the same specifications as in Table 3, except this time using an hourly wage or compensation sample, rather than a weekly one. To save space, we report the results only in percentage terms, which is what most recent studies do, though we caution that these other studies may not convert from log points into percentages in the manner that we do.[12] The hourly results serve

categories of benefits have zero costs for nontrivial numbers of observations. In the online Appendix available with this article at ⟨http://e-jep.org⟩, we describe a methodology that first determines what the average level of benefit costs would be if private sector returns are applied to public sector characteristics and then calculates the ratio of this quantity to the actual average of public sector benefit costs. After controlling for work-level differences, the sectoral differences in health insurance costs are approximately 30–40 percent, and the analogous differences in retirement and savings costs are perhaps 50 percent, as measured relative to the public sector cost.

[12] Relative to simply exponentiating the log differentials, our method of conversion adds 2–4 percentage points in favor of the private sector, depending on the specification.

both as a sensitivity check, and provide an opportunity to reconcile our results with these earlier studies. In the first column, with the Current Population Survey hourly wage sample, we estimate that, with the base controls but without occupational controls, the wages of the private sector exceed those of the state government by 12.4 percent and exceed those of the local government by 11.0 percent, magnitudes which are somewhat lower than their counterparts in the weekly sample. Owing to some differences in sample inclusion criteria and specification and, perhaps, to differences in the method for converting from log points to percentages, our estimates of hourly gaps are wider than those of Schmitt (2010, table 3), who estimated a wage premium of 4–6 percent for the private sector for the same year. The gaps we estimated are about the same as the 11–12 percent wage premium for the private sector estimated by Bender and Heywood (2010) using the 2008 Current Population Survey, although they include "covered by a union contract" as one of their covariates. As noted, we prefer not to control for union status, because we doubt it reflects ability that is portable across sectors. When we experimented with including union coverage in our covariates for the purposes of reconciliation, our gaps widen to 18 to 19 percent in favor of the private sector.

Our results are similar in magnitude to those of Keefe (2010), who estimates that state and local government workers earn 15.57 percent and 9.46 percent below comparable private sector workers, respectively. He uses data from the March Current Population Survey, which compiles earnings on an annual basis. However, he uses organizational size as a control variable, something that is not available in the Merged Outgoing Rotation Group data from the CPS that we use, and a choice that is controversial (Linneman and Wachter 1990). Government is a large employer, and so the inclusion of employer size means that more of the raw differential in favor of government will be viewed as a size effect rather than a government effect, making it more likely that private sector workers will be viewed as overpaid. In the Employer Costs for Employee Compensation data, adding establishment size to our baseline controls shifts differentials by roughly 6 percentage points in favor of the private sector for wages and by about 8 points for compensation, though it is difficult to say whether the effect would be of the same magnitude in the CPS.

The inclusion of occupational indicators in hourly wage regressions moves the relative pay estimates in a direction favorable to government workers, as it did in Table 3, so that state workers are about even with those in the private sector, while local government workers are estimated to earn 5.7 percent more than their private sector counterparts. Recent work on state and local public sector differentials has tended to eschew occupational controls, with the exception of a robustness check by Bender and Heywood (2010). With controls at approximately the same level, they estimate that state workers earned 6.5 percent below private sector workers in 2008, and local workers 3.7 percent below—but, again, they control for union coverage, which may explain part of the difference with our results.

The final two columns of Table 4 display the results for the NCS Employer Costs for Employee Compensation hourly sample, which excludes teachers because of the difficulties in knowing their actual hours worked, as explained earlier. The results for state government hourly wages are not much different than the comparable ones for weekly wages. For local government, the exclusion of teachers and other occupations for which weekly hours are unreliable narrows the raw differential by about 6 percentage points. For the baseline regressions, the hourly and weekly results are also similar, but that is not the case for the regressions with occupational controls, as the estimated edge for local government workers narrows from 9 to 10 percent at both levels of occupational detail to about 7 percent with two-digit controls and 5 percent with detailed controls.

When nonwage benefits are added to hourly wages in the Employer Costs for Employee Compensation microdata in the third column of Table 4, the differentials relative to the second column shift by 10 to 13 percentage points in favor of government workers, a movement that is somewhat greater than the shift in the weekly earnings sample. The hourly compensation differentials are between 3 and 10 percent in favor of state government workers and 13 to 18 percent in favor of local government employees, depending on the specification.

There are two recent studies in this area that include nonwage compensation information, though they do so indirectly. Using unpublished information from the Employer Costs for Employee Compensation for December 2009, Keefe (2010) calculated the ratio of total compensation to wages by major occupation group for state and local government combined and for three establishment-size classes for the private sector. Applying these markups to earnings in the March CPS for 2009, he estimated (in his table 6) that the compensation of a state government employee is 7.55 percent below that of an equivalent private sector employee, while that of a local government employee is 1.84 percent below. Because the same regression specifications for earnings yielded estimated gaps of 15.57 percent and 9.46 percent for state and local government workers, respectively, the implication is that including nonwage compensation moves relative pay estimates in favor of government employees by roughly 8 percentage points.[13]

Bender and Heywood (2010) also use markups from the Employer Costs for Employee Compensation microdata, but with a somewhat different approach than Keefe. Whereas Keefe applied different markups by major occupation group and size to Current Population Survey microdata and then calculated regression-adjusted compensation gaps, Bender and Heywood apply the markups directly to wage gaps that have already been regression adjusted. After having estimated that the average hourly pay in 2000–2008 for state workers was 11.4 percent below that for private sector workers and for local workers was 12.0 percent below, they use markups from 2004–2008 to estimate that state and local compensation remained at 6.8 percent and 7.4 percent below the private sector, respectively.[14] Thus, for

[13] Using Employer Costs for Employee Compensation data for Census divisions rather than for the nation, Keefe also conducted a number of studies for individual states that use compensation markups, available at ⟨http://www.epi.org⟩ (search on "Keefe").

[14] The results reported here use the markup from the entire ECEC. They also apply a markup calculated using just those establishments with 100 or more workers, and obtain correspondingly larger private sector premia.

these authors, controlling for nonwage compensation makes a difference of about 4.5 percentage points.

As noted above, we see a shift of 10 to 13 percentage points in favor of government workers when nonwage compensation is included, which is greater than those movements calculated by other authors. There are at least a few reasons for the difference. Keefe (2010) removes paid leave from compensation before calculating his markups, which raises the relative markup in the private sector. He indicates he takes this step because he believes that paid leave is included in the measure of wages in the Current Population Survey. We have doubts as to whether this is consistently true, but, in any case, paid leave is certainly not included in the Employer Costs for Employee Compensation wage measure. In addition, as we will see, the choice of period for calculating the markups can make a difference, and would have for the Bender and Heywood calculations. Finally, we use a somewhat different set of regression controls and therefore make different comparisons than do these other recent papers.

How Has the Private–Public Pay Gap Changed over Time?

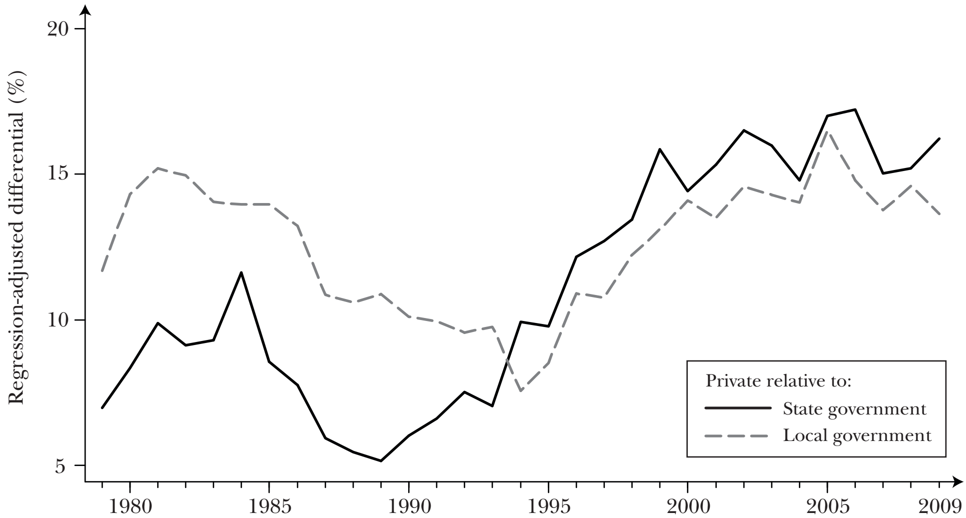
Is the private–public pay gap a new phenomenon? For example, did the gap change during the 2007–09 recession? To get at questions like these, Figure 2 graphs the unexplained differentials for weekly wages (the "base controls" specification in Table 3) for the Current Population Survey data over the 1979–2009 period.[15]

Figure 2 suggests that, for this specification, private sector workers have received higher wages than their state and local counterparts over the entire period. The private sector advantage has ranged from 5 to 17 percent. These series do not appear to suggest cyclical factors as driving mechanisms. During much of the period, from 1989 to 2006, the private sector was gaining on the state sector, though within a relatively narrow range. For the local government sector, the pattern is more of a V-shape, with the private sector losing ground in the first half of the period and regaining it during the second half. When considering wages alone in the CPS, one does not get the impression that there was a recent and sudden surge of the relative pay of the public sector.

Using the Employer Costs for Employee Compensation data, we earlier estimated a current compensation gap in favor of state and local government workers. Ideally, we would like to use the same methodology to estimate differentials over time, but we cannot because the work-level information that is so important for controlling for cross-sectoral differences in the skill distribution of jobs goes back

[15] There were several changes in Current Population Survey data over this period, including a 1994 redesign that altered earnings questions and a lack of allocation flags in certain years. Some of the more important changes involve coding for schooling and occupation. We use the base controls specification to skirt changing CPS occupational coding schemes. The CPS education questions changed in 1992, and we use the approach suggested by Jaeger (1997) to code workers as consistently as possible into five education groups.

Figure 2
Private–Public Percentage Wage Differentials



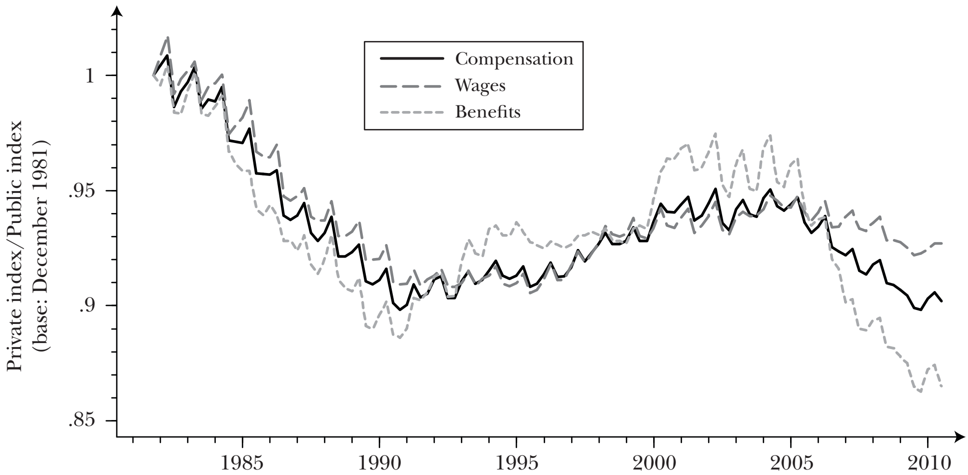
Source: Based on data from the Current Population Survey, Merged Outgoing Rotation Group (CPS–MORG), 1979–2009.
Note: Figure 2 graphs the unexplained differentials for weekly wages (the "base controls" specification in Table 3) for the Current Population Survey data over the 1979–2009 period.

only a few years. Instead, we take an indirect approach, by looking at the relative growth rates of private and public sector compensation, after controlling for changes in employment composition in jobs in each of these sectors.

To do this, we make use of data from the Bureau of Labor Statistics' Employment Cost Index (ECI) series, which are based on the same microdata as the Employer Costs for Employee Compensation.[16] We first take the quarterly private sector ECI indexes for wages, benefits, and total compensation and base each at 100 for December 1981. We base the corresponding public sector ECI indexes in the same manner. Figure 3 plots the ratio of private to public for each series. Faster cost growth in the public sector than in the private manifests as a decline in the plotted series. Because the index growth rates exploit the longitudinal nature of the data at the job level, they abstract from any shifts in the composition of jobs. The series would, however, reflect other developments such as relative wage movements that benefit one sector more than another (for example, a rising return to education) and institutional changes that affect the degree of rent capture in the two sectors. If

[16] The Employment Cost Index is a Laspeyres index for employer costs, with cells defined by industry and occupation. Quarterly rates of change are calculated within each cell using jobs that are in sample both quarters. See Bureau of Labor Statistics (n.d.) for details on the construction of the ECI series. In this figure, we use a combined public sector, combining state and local government. We thank Tom Moehrle for constructing these series, which are not seasonally adjusted.

Figure 3
Changing Private–Public Relative Pay



Source: Based on data from the Bureau of Labor Statistics' Employment Cost Index (ECI) series for 1981–2010.

the methodology of the previous section could be used for this exercise, the former effect would not be present, but the latter would be.

Figure 3 can be divided roughly into three parts: 1) the 1980s, when relative private pay was decreasing (Poterba and Rueben 1994); 2) from roughly 1990 to 2005, when relative pay was fairly stable or, if anything, the private sector was gaining on the public sector; and 3) 2005 to the present, when relative private pay is again falling—though not as fast a rate as in the first period. The third period also differs from the first in that benefits account for a greater portion of total compensation gains (for the public sector relative to the private sector) in the third period than in the first. The wage series in Figures 2 and 3 are broadly consistent, but factoring in benefit costs does seem to change the story somewhat. The relative compensation series shifts 4–5 percent over the last five years. If the analysis on 2009 data in the previous section was possible a half-decade earlier, the findings would likely reflect smaller public–private compensation differentials. Unlike earlier periods, much of the recent shift depends on relative benefit cost changes. Many current popular reports on public sector pay focus on benefits; Figure 3 offers a partial explanation for that focus.

Private–Public Differentials across the Distribution

Up until this point, our focus has been on mean public–private compensation differentials. However, differences in the wage distributions of the public and private sectors extend beyond the differences at the mean—in particular, the

Table 5
Private–Public Percentage Differentials by Quantile

	Quantile				
	10th	*25th*	*median*	*75th*	*90th*
CPS weekly earnings					
Private–State Government	–4.5**	–1.8*	2.9**	7.9**	11.7**
Private–Local Government	–10.5**	–8.4**	–2.9**	3.4**	9.0**
ECEC weekly earnings					
Private–State Government	0.9	0.5	3.5*	5.6**	7.5**
Private–Local Government	–7.0**	–7.5**	–5.6**	–4.8**	–0.7
ECEC weekly compensation					
Private–State Government	–13.2**	–11.6**	–9.3**	–7.5**	–5.5**
Private–Local Government	–19.7**	–18.7**	–16.4**	–14.0**	–10.5**

Sources: Based on data from the 2009 Current Population Survey, Merged Outgoing Rotation Group (CPS–MORG) file and data from the Employer Costs for Employee Compensation (ECEC) sample in the 2009 National Compensation Survey (NCS).
Notes: Columns report quantile regression results at the given percentile. Controls are the same as for the "base controls and major occupation" weekly earnings models of Table 3.
** and * indicate statistical significance at the 1 and 5 percent levels, respectively.

distribution of pay in the public sector tends to be less dispersed (Poterba and Rueben 1994; Belman and Heywood 2004a). We revisit this point here, using quantile regression techniques to assess public–private differentials at different points of the distribution.

The dummy variable approach uses the single regression equation:

$$\ln(y_{jk}) = \mathbf{X}_{jk}\boldsymbol{\beta} + S_k\delta_s + L_k\delta_l + \varepsilon_{jk}.$$

Again, for sector j, which can be state, local, or private, and individual k, y_{jk} is either a wage or compensation measure, $\mathbf{X}_{jk}$ is a vector of characteristics, and $\boldsymbol{\beta}$ is the corresponding coefficient vector, which does not differ by sector. This specification also includes S_k and L_k, which are dummy variables for state and local government with corresponding coefficients δ_s and δ_l. In this case, the coefficients on the two sectoral dummy variables directly provide the log differentials between state government pay relative to private pay (the omitted variable), and between local government pay and private pay. Table 5 gives estimates of this specification, with major occupation controls, using quantile regressions. Since we are estimating differentials at particular points in the distribution, here a simple exponentiation of the public sector coefficients is appropriate for deriving percentage differences, and we present these estimates in Table 5.

In the Current Population Survey data in the first row of Table 5, consistent with expectations, the private sector premium rises as one moves across the columns, through the distribution of weekly earnings. Below the median, the private sector premium is negative (higher relative pay for state and local government), but, at the 90th percentile, the private sector premium is 11.7 percent relative to state government and 9.0 percent versus local government.

In the rows of Table 5 based on the Employer Costs for Employee Compensation earnings and compensation microdata, the same movement in favor of the private sector is evident as one moves up the quantiles, but the interquantile spread is narrower in the ECEC than in the data from the Current Population Survey. In the CPS, there is a spread of 16.2 percentage points between private–public percentage differentials in the 10th and 90th percentiles for the state government and one of 19.5 points for the local government. In the ECEC, the private sector premium for weekly earnings relative to state workers is 0.9 percent at the 10th percentile versus 7.5 percent at the 90th percentile, a difference of only 6.6 percentage points. The spread across the quantiles in the bottom row (ECEC weekly compensation, private–local government) is a somewhat larger 9.2 percentage points, yet even at the 90th percentile, compensation for private sector workers is lower than that for local government workers, by 10.5 percent.

Concluding Remarks

We have sought to address the broad question of whether workers in state and local government are better compensated than their private sector counterparts. A more detailed analysis might delve into particular occupations. For example, teachers are a large part of the local government workforce, but assessing teacher pay is a difficult task because private and public sector teachers operate in different environments. We have not tried to take into account the unique dangers faced by certain public workers like firefighters and police officers. More generally, implicit in some of the discussion in the popular press is the question of whether state and local government jobs are *better* jobs than those in the private sector, which would require considering not only pecuniary benefits, but also nonpecuniary ones. For example, job security is better in the public sector, which has been an especially salient point in recent years. According to data from the Job Openings and Labor Turnover Survey (JOLTS), the annual layoff and discharge rate for the private sector ranged from 17.6 percent to 22.8 percent in 2006–2010 in contrast to a range of 5.9 percent to 7.0 percent for state and local government. Other nonpecuniary benefits that could affect the relative attractiveness of public sector employment might involve working conditions. Work-related injury rates appear to be higher in the public sector (Bureau of Labor Statistics 2011), and there may be differences in work effort or work schedules for which our regression controls do not account completely. An analysis of nonpecuniary job benefits might use job queues (Krueger 1988a, b; Heywood and Mohanty 1993), where higher numbers of applicants may be indicative of the presence of rents for a job, or might try to directly price out the value of a benefit (Richwine and Biggs 2011), but such analyses are beyond the scope of this paper.

Are state and local government workers better paid than similar workers in the private sector? When considering wages only, the answer is ambiguous. When we add nonwage compensation, however, public sector workers do appear to be better compensated, although the magnitude of the difference depends upon which sector of government is being considered and the degree to which occupations are controlled for. With no controls for occupations, we estimate that compensation in state government is higher than in the private sector by 3.2 percent in the weekly sample and 3.7 percent in an hourly sample. Local government workers are even more highly compensated, with the differential being 10.5 percent in the weekly sample and 12.9 percent in the hourly sample. The addition of occupation controls, particularly those for two-digit occupations, serves to widen the premium for the government. We should note, however, that none of these results control for establishment size. Such a control, in models with detailed occupation, tends to reduce the private–local government compensation differential, and approximately equalizes the compensation of state and private sector workers. There has been no major recent rise in the relative wages of state and local government workers, but recent years have seen a faster rate of increase in benefit costs in the public sector, translating into a somewhat faster increase in overall compensation costs.

Future work might seek to reconcile the raw wage differentials in the Current Population Survey and the Employer Costs for Employee Compensation data. One avenue would be to attempt replication of the ECEC wage findings in other administrative records or establishment survey data. Other avenues are to identify important survey measurement errors, or to determine whether the CPS and ECEC substantially measure different things (and if so, what the preferred construct is).

Finally, we wonder about the forces driving the premia we estimate. There are interesting agency questions here: politicians are agents for citizens in compensation settings and may face conflicts of interest in dealing with the public sector workforce. There are likely to be additional agency considerations involving local government pay setting, since wages are determined by local government actors and those wage choices can affect state-level liabilities in benefit funding. There are a wide variety of state- and local-level political institutions, including different rules governing bargaining, different budgets and expenditures, and different abilities to renege on promised benefits, and it would be interesting to ascertain how such structural differences influence observed outcomes.

■ *The views expressed here are those of the authors and do not necessarily reflect the views or policies of the Bureau of Labor Statistics (BLS) or any other agency of the U.S. Department of Labor. We thank participants at a BLS seminar for useful discussions. We are particularly grateful to David Autor and Timothy Taylor for their guidance and to Will Carrington, Chad Jones, Chinhui Juhn, John List, and Rick Schumann for helpful comments.*

References

Allegretto, Sylvia A., Sean P. Corcoran, and Lawrence Mishel. 2004. *How Does Teacher Pay Compare? Methodological Challenges and Answers.* Washington, DC: Economic Policy Institute.

Belman, Dale, and John S. Heywood. 2004a. "Public-Sector Wage Comparability: The Role of Earnings Dispersion." *Public Finance Review* 32(6): 567–87.

Belman, Dale, and John S. Heywood. 2004b. "Public Wage Differentials and the Treatment of Occupational Differences." *Journal of Policy Analysis and Management* 23(1): 135–52.

Bender, Keith A., and John S. Heywood. 2010. "Out of Balance? Comparing Public and Private Sector Compensation over 20 Years." Report for the National Institute for Retirement Security and the Center for State and Local Government Excellence, April.

Blackburn, McKinley L. 2008. "Are Union Wage Differentials in the United States Falling?" *Industrial Relations* 47(3): 390–418.

Blinder, Alan. 1973. "Wage Discrimination: Reduced Form and Structural Estimates." *Journal of Human Resources* 8(4): 436–55.

Brown, Charles C., and James L. Medoff. 1988. "Employer Size, Pay, and the Ability to Pay in the Public Sector." In *When Public Sector Workers Unionize,* edited by Richard B. Freeman and Casey Ichniowski, 195–216. Chicago: University of Chicago Press.

Bureau of Labor Statistics, U.S. Department of Labor. 2003. *National Compensation Survey: Guide for Evaluating Your Firm's Jobs and Pay.* http://www.bls.gov/ncs/ocs/sp/ncbr0004.pdf.

Bureau of Labor Statistics, U.S. Department of Labor. 2011. "Workplace Injuries and Illnesses—2010." News Release, October 20. http://www.bls.gov/news.release/pdf/osh.pdf.

Bureau of Labor Statistics, U.S. Department of Labor. N.d. "National Compensation Measures." Chap. 8 in *Handbook of Methods.* http://www.bls.gov/opub/hom/pdf/homch8.pdf.

Duan, Naihua. 1983. "Smearing Estimate: A Nonparametric Retransformation Method." *Journal of the American Statistical Association* 78(383): 605–10.

Famulari, Melissa, and Marilyn E. Manser. 1989. "Employer-Provided Benefits: Employer Costs vs. Employee Value." *Monthly Labor Review* 112(12): 24–32.

Fortin, Nicole, Thomas Lemieux, and Sergio Firpo. 2010. "Decomposition Methods in Economics." NBER Working Paper 16045.

Gittleman, Maury, and Brooks Pierce. 2007. "New Estimates of Union Wage Effects." *Economics Letters* 95(2): 198–202.

Gittleman, Maury, and Brooks Pierce. 2011. "Inter-Industry Wage Differentials, Job Content and Unobserved Ability." *Industrial and Labor Relations Review* 64(2): 356–74.

Gregory, Robert G., and Jeff Borland. 1999. "Recent Developments in Public Sector Labor Markets." In *Handbook of Labor Economics, Vol. 3,* edited by Orley C. Ashenfelter and David Card, 3574–3630. New York: Elsevier Science.

Gyourko, Joseph, and Joseph Tracy. 1988. "An Analysis of Public- and Private-Sector Wages Allowing for Endogenous Choices of Both Government and Union Status." *Journal of Labor Economics* 6(2): 229–53.

Heywood, John S., and Madhu S. Mohanty. 1993. "Testing for State and Local Government Job Queues." *Journal of Labor Research* 14(4): 455–67.

Hirsch, Barry T., and Edward J. Schumacher. 2004. "Match Bias in Wage Gap Estimates due to Earnings Imputation." *Journal of Labor Economics* 22(3): 689–722.

Hirsch, Barry T., Michael L. Wachter, and James W. Gillula. 1999. "Postal Service Compensation Standard." In *Research in Labor Economics,* Vol. 18, edited by S. W. Polachek and J. Robst, 243–79. JAI Press.

Jaeger, David A. 1997. "Reconciling the Old and New Census Bureau Education Questions: Recommendations for Researchers." *Journal of Business and Economics Statistics* 15(3): 300–309.

Keefe, Jeffrey. 2010. "Debunking the Myth of the Overcompensated Public Employee: The Evidence." EPI Briefing Paper 276, Economic Policy Institute. September 15.

Krueger, Alan B. 1988a. "Are Public Sector Workers Paid More than Their Alternative Wage?" In *When Public Sector Workers Unionize,* edited by Richard B. Freeman and Casey Ichniowski, 217–42. Chicago: University of Chicago Press.

Krueger, Alan. B. 1988b. "The Determinants of Queues for Federal Jobs." *Industrial and Labor Relations Review* 41(4): 567–81.

Lee, Sang-Hyop. 2004. "A Reexamination of Public-Sector Wage Differentials in the United States: Evidence from the NLSY with Geocode." *Industrial Relations* 43(2): 448–72.

Lemieux, Thomas. 2010. "What Do We Really Know about Changes in Wage Inequality?" In *Labor in the New Economy,* edited by Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, 42–59. Chicago: University of Chicago Press.

Levenson, Alec, and Cindy Zoghi. 2011.

"Occupations, Human Capital and Skills." *Journal of Labor Research* 31(4): 365–86.

Linneman, Peter D., and Michael L. Wachter. 1990. "The Economics of Federal Compensation." *Industrial Relations* 29(1): 58–76.

Manning, Willard G., N. Duan, and W. H. Rogers. 1987. "Monte Carlo Evidence on the Choice between Sample Selection and Two-Part Models." *Journal of Econometrics* 35(1): 59–82.

Moulton, Brent R. 1990. "A Reexamination of the Federal–Private Wage Differential in the United States." *Journal of Labor Economics* 8(2): 270–93.

Oaxaca, Ronald. 1973. "Male–Female Wage Differentials in Urban Labor Markets." *International Economic Review* 14(3): 693–709.

Park, Jin Heum. 1994. "Estimation of Sheepskin Effects and Returns to Schooling Using the Old and the New CPS Measures of Educational Attainment." Princeton University, Industrial Relations Section Working Paper 338.

Pierce, Brooks. 2001. "Compensation Inequality." *Quarterly Journal of Economics* 116(4): 1493–1525.

Pierce, Brooks. 2010. "Recent Trends in Compensation Inequality." In *Labor in the New Economy,* edited by Katharine G. Abraham, James R. Spletzer, and Michael J. Harper, 63–98. Chicago: University of Chicago Press.

Poterba, James M., and Kim S. Rueben. 1994. "The Distribution of Public Sector Wage Premia: New Evidence Using Quantile Methods." NBER Working Paper 4734.

Podgursky, Michael, and Ruttaya Tongrut. 2006. "(Mis-)Measuring the Relative Pay of Public School Teachers." *Education Finance and Policy* 1(4): 425–40.

President's Pay Agent. 2009. "Report on Locality-Based Comparability Payments for the General Schedule." Annual Report of the President's Pay Agent, 2009. http://www.opm.gov/oca/payagent/2009/2009PayAgentReport.pdf.

Richwine, Jason, and Andrew Biggs. 2011. "Are California Public Employees Overpaid?" A Report of the Heritage Center for Data Analysis, CDA 11-01.

Schmitt, John. 2010. "The Wage Penalty for State and Local Government Employees." Center for Economic and Policy Research, May. http://www.cepr.net/documents/publications/wage-penalty-2010-05.pdf.

Schumann, Richard. 2008. "Work Schedules in the National Compensation Survey." Originally posted July 28 at U.S. Department of Labor, Bureau of Labor Statistics website. http://www.bls.gov/opub/cwc/cm20080722ar01p1.htm.

Troske, Kenneth R. 1999. "Evidence on the Employer Size-Wage Premium from Worker-Establishment Matched Data." *Review of Economics and Statistics* 81(1): 15–26.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 243–250

Recommendations for Further Reading

Timothy Taylor

This section will list readings that may be especially useful to teachers of undergraduate economics, as well as other articles that are of broader cultural interest. In general, with occasional exceptions, the articles chosen will be expository or integrative and not focus on original research. If you write or read an appropriate article, please send a copy of the article (and possibly a few sentences describing it) to Timothy Taylor, preferably by e-mail at ⟨taylort@macalester.edu⟩, or c/o *Journal of Economic Perspectives*, Macalester College, 1600 Grand Ave., Saint Paul, Minnesota, 55105.

Smorgasbord

W. Brian Arthur discusses the scope of "The Second Economy." "I want to argue that something deep is going on with information technology, something that goes well beyond the use of computers, social media, and commerce on the Internet. Business processes that once took place among human beings are now being executed electronically. They are taking place in an unseen domain that is strictly digital. On the surface, this shift doesn't seem particularly consequential—it's almost something we take for granted. But I believe it is causing a revolution no less important and dramatic than that of the railroads. It is quietly creating a second economy, a digital one. . . . Now this second, digital economy isn't producing anything tangible. It's not making my bed in a hotel, or bringing me orange juice in the morning. But it is running an awful lot of the economy. It's helping architects

■ *Timothy Taylor is Managing Editor,* Journal of Economic Perspectives*, based at Macalester College, Saint Paul, Minnesota. His e-mail address is ⟨taylort@macalester.edu⟩ and he blogs at ⟨http://conversableeconomist.blogspot.com/⟩.*

doi=10.1257/jep.26.1.243

design buildings, it's tracking sales and inventory, getting goods from here to there, executing trades and banking operations, controlling manufacturing equipment, making design calculations, billing clients, navigating aircraft, helping diagnose patients, and guiding laparoscopic surgeries. Such operations grow slowly and take time to form. . . . Is this the biggest change since the Industrial Revolution? Well, without sticking my neck out too much, I believe so. In fact, I think it may well be the biggest change ever in the economy. It is a deep qualitative change that is bringing intelligent, automatic response to the economy. There's no upper limit to this, no place where it has to end. . . . I think that for the rest of this century, barring wars and pestilence, a lot of the story will be the building out of this second economy, an unseen underground economy that basically is giving us intelligent reactions to what we do above the ground." *McKinsey Quarterly*. October 2011. Free registration needed, at ⟨http://www.mckinseyquarterly.com/Strategy/Growth/The_second_economy_2853⟩.

Choices, published by the Agricultural and Applied Economics Association, has a set of six short readable articles with diverse views on the subject: "Should Soft Drinks Be Taxed More Heavily?" For example, Jason Fletcher writes: "[S]oft drink consumption has increased by almost 500% in the past 50 years, and recent data suggest it represents 7% of overall energy intake in adults and often larger proportions in children . . . a 16% share of calories in youth ages 12–19 and 11% in children ages 2–11. . . . We know that soda consumption is an important share of total consumption, and ample evidence suggests that maintained reductions in consumption of approximately 100 calories per day—less than a can of soda—could halt weight gain for 90% of the population . . . Fletcher continues: "[W]hile individuals in states with higher soda taxes have lower soda consumption, these individuals completely offset the reductions in calories from soda by consuming other high-calorie beverages, such as milk and juice." October 2011. At ⟨http://www.choicesmagazine.org/choices-magazine/policy-issues/should-soft-drinks-be-taxed-more-heavily⟩.

Zsolt Darvas tells "A Tale of Three Countries: Recovery after Banking Crisis." "Three small, open European economies—Iceland, Ireland and Latvia with populations of 0.3, 4.4 and 2.3 million respectively—got into serious trouble during the global financial crisis. Behind their problems were rapid credit growth and expansion of other banking activities in the years leading up to the crisis, largely financed by international borrowing. This led to sharp increases in gross (Iceland and Ireland) and net (Iceland and Latvia) foreign liabilities. Credit booms fuelled property-price booms and a rapid increase in the contribution of the construction sector to output—above 10 percent in all three countries. While savings–investment imbalances in the years of high growth were largely of private origin, public spending kept up with the revenue overperformance that was the consequence of buoyant economic activity. During the crisis, property prices collapsed, construction activity contracted and public revenues fell, especially those related to the previously booming sectors. . . . [T]he crisis hit Latvia harder than any other country, and Ireland also suffered heavily, while Iceland exited the crisis with the smallest fall in employment, despite the greatest shock to the financial system. . . . There were marked differences in policy mix: currency collapse in Iceland but not in Latvia, letting banks fail in Iceland but not in Ireland, and the introduction of strict capital controls only in Iceland. The speed of fiscal consolidation was fastest in Latvia and slowest in Ireland. Recovery has started in all three countries. Iceland seems to have the right policy mix." *Bruegel Policy Contribution*, December 2011, Issue 2011/19. At ⟨http://www.bruegel.org/download/parent/663-a-tale-of-three-countries-recovery-after-banking-crises/file/1534-a-tale-of-three-countries-recovery-after-banking-crises/⟩.

The Committee on Economic and Environmental Impacts of Increasing Biofuels Production of the National Research Council has published "Renewable Fuel Standard: Potential Economic and Environmental Effects of U.S. Biofuel Policy." Some findings: "Only in an economic environment characterized by high oil prices, technological breakthroughs, and a high implicit or actual carbon price would biofuels be cost-competitive with petroleum-based fuels." "RFS2 [renewable fuel standards] may be an ineffective policy for reducing global GHG [greenhouse gas] emissions because the effect of biofuels on GHG emissions depends on how the biofuels are produced and what land-use or land-cover changes occur in the process." "Absent major technological innovation or policy changes, the RFS2-mandated consumption of 16 billion gallons of ethanol-equivalent cellulosic biofuels is unlikely to be met in 2022." A "prepublication copy" can be downloaded (with free registration) at ⟨http://www.nap.edu/catalog.php?record_id=13105⟩.

Kathleen Short describes "The Research Supplemental Poverty Measure: 2010." Here's a table from that report summarizing how the new supplemental measure of poverty from the U.S. Census Bureau differs from the official measure:

	Official poverty measure	*Supplemental poverty measure*
Measurement units	Families and unrelated individuals	All related individuals who live at the same address, including any coresident unrelated children who are cared for by the family (such as foster children) and any cohabitors and their children
Poverty threshold	Three times the cost of minimum food diet in 1963	The 33rd percentile of expenditures on food, clothing, shelter, and utilities (FCSU) of consumer units with exactly two children multiplied by 1.2
Threshold adjustments	Vary by family size, composition, and age of householder	Geographic adjustments for differences in housing costs and a three parameter equivalence scale for family size and composition
Updating thresholds	Consumer Price Index: all items	Five year moving average of expenditures on FCSU
Resource measure	Gross before-tax cash income	Sum of cash income, plus in-kind benefits that families can use to meet their FCSU needs, minus taxes (or plus tax credits), minus work expenses, minus out-of-pocket medical expenses

Current Population Reports P60-241. November 2011. At ⟨http://www.census.gov

/hhes/povmeas/methodology/supplemental/research/Short_ResearchSPM 2010.pdf⟩.

Thorstein Beck has edited an e-book on *The Future of Banking*. It includes 12 readable essays by expert economists, based on their academic research. As one example, here is Neeltje van Horen on "The Changing Role of Emerging-Market Banks." "Although many in the West are not familiar with emerging-market banks, they are by no means small. In fact, the world's biggest bank in market value is China's ICBC. The global top 25 includes eight emerging-market banks. Among these, three other Chinese banks (China Construction Bank, Agricultural Bank of China, and Bank of China), three Brazilian banks (Itaú Unibanco, Banco do Brasil, and Banco Bradesco) and one Russian bank (Sberbank). While excess optimism might have inflated these market values, these banks are large with respect to other measures as well. In terms of assets all these banks are in the top 75 worldwide, with all four Chinese banks in the top 20. Furthermore, in 2010 emerging-market banks as a group accounted for roughly 30% of global profits, a third of global revenues, and half of tier 1 capital." She points out that these banks have in some ways been sheltered from the financial turmoil of the last few years, are located in fast-growing countries with high domestic savings rates, and thus are poised to continue rapid growth. Vox. October 25, 2011. At ⟨http://www.voxeu.org/index.php?q=node/7147⟩.

From the Federal Reserve

Jun Nie and Ethan Struby ask: "Would Active Labor Market Policies Help Combat High U.S. Unemployment?" By "active" labor market policies, they mean government support for job training, job search, incentives for private firms to hire, or even direct job creation. In contrast, "passive" policies are unemployment benefits or early retirement. "The level of spending on labor market policies differs widely across OECD countries. Between 1998 and 2008 in 21 OECD countries, total expenditures on passive and active labor market policies as a fraction of GDP ranged from about 4 percent in Denmark to 0.25 percent in the United Kingdom (Chart 3). The United States is near the bottom of this list, spending slightly less than 0.5 percent of GDP on labor market policies during this time. In addition, the fraction of spending on active versus passive policies differs across countries. Outside the United States, the average country in Chart 3 devoted 59 percent of labor market policy expenditures to PLMP [passive labor market policies] and 41 percent to ALMP [active labor market policies]. In the United States, however, 70 percent of expenditures went to PLMP and 30 percent went to ALMP." They present some evidence supporting greater spending assistance for job training and for job search. *Economic Review,* Federal Reserve Bank of Kansas City, Third Quarter 2011, pp. 35–69. At ⟨http://www.kansascityfed.org/publicat/econrev/pdf/11q3Nie-Struby.pdf⟩.

Christopher J. Neely tells of the international exchange rate intervention to stabilize the value of the Japanese yen in the aftermath of the March 2011 earthquake in "A Foreign Exchange Intervention in an Era of Restraint." Neely reports that the major central banks of the world have carried out only three exchange rate interventions since 1995: an intervention after Japan's quake in March 2011, an intervention soon after the start of the euro in September 2000, and an intervention in the yen after East Asia's financial crisis in 1998. He tells what happened in each case, and sums up: "Since 1995 most advanced governments/central banks have used intervention only very sparingly as a policy tool. Examination of coordinated interventions during this period shows that intervention is not a magic wand that authorities can use to move exchange rates at will. It can be a very effective tool in certain circumstances, however, to coordinate market expectations about fundamental values of the exchange rate and calm disorderly foreign exchange markets by reintroducing two-sided risk." *Review,* Federal Reserve Bank of St. Louis, September/October 2011, pp. 303–324. At ⟨http://research.stlouisfed.org/publications/review/11/09/303-324Neely.pdf⟩.

Inequality around the World

The OECD has published *Divided We Stand: Why Inequality Keeps Rising.* "In OECD countries today, the average income of the richest 10% of the population is about nine times that of the poorest 10%—a ratio of 9 to 1. However, the ratio varies widely from one country to another. It is much lower than the OECD average in the Nordic and many continental European countries, but reaches 10 to 1 in Italy, Japan, Korea, and the United Kingdom; around 14 to 1 in Israel, Turkey, and the United States; and 27 to 1 in Mexico and Chile. "Benefits had a much stronger impact on inequality than the other main instruments of cash distribution—social contributions or taxes. . . . The most important benefit-related determining factor in overall distribution, however, was not benefit levels but the number of people entitled to transfers." "However, redistribution strategies based on government transfers and taxes alone would be neither effective nor financially sustainable. First, there may be counterproductive disincentive effects if benefit and tax reforms are not well designed. Second, most OECD countries currently operate under a reduced fiscal space which exerts strong pressure to curb public social spending and raise taxes. Growing employment may contribute to sustainable cuts in income inequality, provided the employment gains occur in jobs that offer career prospects. Policies for more and better jobs are more important than ever." December 2011. The report can be read for free online, although the software for doing so is a bit awkward, and an overview chapter can be downloaded as a PDF, at ⟨http://www.oecd.org/document/51/0,3746,en_2649_33933_49147827_1_1_1_1,00.html⟩.

The September 2011 issue of *Finance & Development* has four articles about economic inequality around the world. For example, Branko Milanovic writes "More or Less: Income Inequality Has Risen over the Past Quarter-Century Instead of Falling as Expected." "The view that income inequality harms growth—or that improved equality can help sustain growth—has become more widely held in recent years. . . . Historically, the reverse position—that inequality is good for growth—held sway among economists. The main reason for this shift is the increasing importance of human capital in development. When physical capital mattered most, savings and investments were key. Then it was important to have a large contingent of rich people who could save a greater proportion of their income than the poor and invest it in physical capital. But now that human capital is scarcer than machines, widespread education has become the secret to growth. And broadly accessible education is difficult to achieve unless a society has a relatively even income distribution. Moreover, widespread education not only demands relatively even income distribution but, in a virtuous circle, reproduces it as it reduces income gaps between skilled and unskilled labor. So economists today are more critical of inequality than they were in the past." At ⟨http://www.imf.org/external/pubs/ft/fandd/2011/09/⟩.

The *World Development Report 2012* from the World Bank is centered on the theme: "Gender Equality and Development." The report suggests considerable worldwide progress in gender equality in education and health. "Although boys are more likely than girls to be enrolled in primary school, girls make better progress—lower repetition and lower dropout rates—than boys in all developing regions. . . . Gender now explains very little of the remaining inequality in school enrollment . . ." "In most world regions, life expectancy for both men and women has consistently risen, with women on average living longer than men . . . On various other aspects of health status and health care, differences by sex are small. In many low-income countries, the proportion of children stunted, wasted, or underweight remains high, but girls are no worse off than boys. . . . Similarly, there is little evidence of systematic gender discrimination in the use of health services or in health spending." The report also points out where a high degree of gender inequality persists: for example, lack of female participation in certain occupations and in political leadership. Also, gender bias at birth remains strong in many places: "In China and India, sex ratios at birth point to a heavily skewed pattern in favor of boys. Where parents continue to favor sons over daughters, a gender bias in sex-selective abortions, female infanticide, and neglect is believed to account for millions of missing girls at birth. In 2008 alone, an estimated 1 million girls in China and 250,000 girls in India were missing at birth." Available by searching at WorldBank.org.

About Economists

Douglas Clement has a wide-ranging "Interview with Daron Acemoglu" in *The Region* (Federal Reserve Bank of Minneapolis). On the Dodd–Frank financial reform legislation: "I think the problem with the Dodd–Frank Act is that the amount of good it contains seems to be dwarfed by the amount of additional minute details it contains. That fails to achieve the intent of the regulation. It also gives better regulation a bad name, because people who are opposed to regulation can easily point to the page after page after page of paperwork and procedural things that Dodd–Frank wants you to do. And I am not convinced that the Dodd-Frank Act is going to prevent the next financial collapse if the financial system actually continues on its current trajectory." On economic growth and political institutions: "But later in college and graduate school, I started working on issues related to human capital, economic growth and so on. But then after a while, I sort of realized, well, you know, the real problems of economic growth aren't just that some countries are technologically innovative and some aren't, and some countries have high savings rates and some don't. They are really related to the fact that societies have chosen radically different ways of organizing themselves. So there is much meaningful heterogeneity related to economic outcomes *in the political structures* of societies." September 2011, pp. 18–31. At ⟨http://www.minneapolisfed.org/publications_papers/pub_display.cfm?id=4733⟩.

George S. Tavlas discusses the nineteenth century American economist Alexander Del Mar in "The Money Man." James Tobin called him "one of the most important U.S. monetary economists of the 19th century"; Robert Mundell called him "too hot to handle." Del Mar was a co-founder in 1865 of the *New York Social Science Review: Devoted to Political Economy and Statistics*, often thought of as one of the first economics journals published in the United States. He was also the first director of the U.S. Bureau of Statistics in 1866—which later evolved into the U.S. Department of Commerce. Del Mar was far ahead of his time in thinking of money as a unit of account and a yardstick of value. In 1886, he proposed that the government should commit to increasing the money supply by 3 percent annually—thus scooping Milton Friedman's similar proposal by about seven decades. *American Interest*, November–December 2011, pp. 110–14. At ⟨http://www.the-american-interest.com/article.cfm?piece=1112⟩.

Discussion Starters

Jeffrey Frankel has a readable overview of the arguments over "The Curse: Why Natural Resources are Not Always a Good Thing." "It is striking how often countries that are rich with oil, minerals or fertile land have failed to grow more rapidly than those without. Angola, Nigeria and Sudan are all awash in petroleum, yet most of their citizens are bitterly poor. Meanwhile, East Asian economies, including Japan, Korea, Taiwan, Singapore and Hong Kong, have achieved Western-level standards of living despite being rocky islands (or peninsulas) with virtually no exportable natural resources. This is the phenomenon known to economists as the 'natural resources curse.' The evidence for its existence is more than anecdotal. The curse shows up in econometric tests of the determinants of economic performance across a comprehensive sample of countries." Before suggesting some policy responses, Frankel reviews five possible reasons behind the "curse": 1) Commodity prices fluctuate a lot, so an economy that depends on commodity exports will be hit by a series of shocks; 2) An economy focused on natural resources diverts land, labor, and capital from other sectors of the economy, like manufacturing; 3) Natural resource endowments can foster corruption and weak institutions, as different groups jostle for control of the income from the resources; 4) High exports of natural resources can lead to currency appreciation which then disadvantages all other exports; and 5) Natural resources can be depleted. *Milken Institute Review*, Fourth Quarter 2011. Available (with free registration) at ⟨http://www.milkeninstitute.org/publications/⟩.

Ahmad Faruqui and Jennifer Palmer look at how households react to variable pricing of electricity in "Dynamic Pricing and Its Discontents." "[A]lmost all analyses of pilot results show that customers do respond to dynamic pricing rates by lowering peak usage. Indeed, in 24 different pilots involving a total of 109 different tests of time-varying rates—covering many different locations, time periods, and rate designs—customers have reduced peak load on dynamic rates relative to flat rates, with a median peak reduction (or demand response) of 12 percent. . . . In other words, the demand for electricity does respond to price, just like the demand for other products and services that consumers buy." "At the national level, an assessment carried out for FERC [Federal Energy Regulatory Commission] two years ago showed that the universal application of dynamic pricing in the United States had the potential for quintupling the share of U.S. peak demand that could be lowered through demand response, from 4 percent to 20 percent. Another assessment quantified the value of demand response and showed that even a 5 percent reduction in U.S. peak demand could lower energy costs $3 billion a year." *Regulation*, Fall 2011, pp. 16–22. At ⟨http://www.cato.org/pubs/regulation/regv34n3/regv34n3-5.pdf⟩. This paper is a useful complement to the paper by Paul Joskow in this issue.

The U.S. Government Accountability Office reported on: "Horse Welfare: Action Needed to Address Unintended Consequences from Cessation of Domestic Slaughter." "Since fiscal year 2006, Congress has annually prohibited the use of federal funds to inspect horses destined for food, effectively prohibiting domestic slaughter. . . . Since domestic horse slaughter ceased in 2007, the slaughter horse market has shifted to Canada and Mexico. From 2006 through 2010, U.S. horse exports for slaughter increased by 148 and 660 percent to Canada and Mexico, respectively. As a result, nearly the same number of U.S. horses was transported to Canada and Mexico for slaughter in 2010—nearly 138,000—as was slaughtered before domestic slaughter ceased. . . . GAO analysis of horse sale data estimates that closing domestic horse slaughtering facilities significantly and negatively affected lower-to-medium priced horses by 8 to 21 percent; higher-priced horses appear not to have lost value for that reason. . . . Comprehensive, national data are lacking, but state, local government, and animal welfare organizations report a rise in investigations for horse neglect and more abandoned horses since 2007." In November 2011, President Obama signed a new agriculture bill into law that will probably allow U.S. horse slaughter facilities to re-open. June 2011. At ⟨http://www.gao.gov/new.items/d11228.pdf⟩.

Journal of Economic Perspectives—Volume 26, Number 1—Winter 2012—Pages 251–252

Notes

For additional announcements, check out the continuously updated* JEP *online Bulletin Board, *⟨http://www.aeaweb.org/bulletinboard.php⟩. Calls for papers, notices of professional meetings, and other announcements of interest to economists should be submitted to Ann Norman at ⟨jep@jepjournal.org⟩ in one or two paragraphs containing the relevant information. These will be posted at the* JEP *online Bulletin Board. Given sufficient lead time, we will also print a shorter, one-paragraph version of your notice in the "Notes" section of the* Journal of Economic Perspectives. ***Deadlines for "Notes":*** *March 20 for the* JEP *Spring issue, which mails the end of May; June 20 for the* JEP *Summer issue, which mails the end of August; September 20 for the* JEP *Fall issue, which mails the end of November; and December 10 for the* JEP *Winter issue, which mails the end of February. We reserve the right to edit material received.*

Call for Sessions and Papers for the January 2013 American Economic Association Annual Meeting. Members wishing to give papers or organize complete sessions for the program for the AEA meetings in San Diego, January 4–6, 2013, are invited to submit proposals electronically (NO hardcopies!) to Professor Claudia Goldin via the American Economic Association web site at ⟨http://www.aeaweb.org/Annual_Meeting/submissions.php⟩. The submission portal for the 2013 annual AEA meeting opens on March 1, 2012. While papers covering a wide array of topics in economics will be included on the 2013 program, we are especially interested in receiving proposals that interpret current economic issues in light of the past.

Econ-Harmony allows prospective 2013 AEA Meetings individual paper submitters who are members of the Association to post information about their paper and search for others with similar interests who might join them to form a complete session submission, and provides an opportunity to volunteer as a session chair. Of papers listed on Econ-Harmony for the 2010 and 2011 Meetings, 112 were submitted to the AEA program as individual papers, and 91 were submitted as part of a complete session. Of those submitted as part of a complete session, 15% made the program; of those submitted individually, 9% made the program. The Program Committee had no knowledge of which papers had been listed on Econ-Harmony when they decided which papers and sessions made the program. Thirty-one percent of 273 submitted complete sessions and 13% of 1404 submitted individual papers made it onto the 2010 Program; 39% of the 287 submitted complete sessions and 17% of 897 submitted individual papers made it onto the 2011 AEA Annual Meeting Program. Econ-Harmony will open Feburary 12, 2012. Find it at ⟨http://www.aeaweb.org/econ-harmony/⟩.

The Committee on Economic Education (CEE) announces its second annual conference on teaching undergraduate and graduate economics, and research on economic education at all levels (including precollege). The conference, cosponsored by the *Journal of Economic Education*, will be at the Royal Sonesta Hotel in Boston, May 30–June 1, 2012, with a dinner at the Boston Fed. Plenary speakers include:

- Daron Acemoglu, "Incorporating Long-run Economic Development into the Undergraduate Economics Curriculum"
- Susan Athey, "Economics Education for the Internet Age: Design, Analysis, and Experimentation in Large-Scale Online Marketplaces"
- Peter Diamond, "Unemployment, Vacancies, Wages"
- Jeff Fuhrer, "Monetary Policy in the Wake of the Great Recession"
- Greg Mankiw, "Recent Challenges Facing Monetary and Fiscal Policy, and What They Mean for What We Teach"

Concurrent sessions will also be scheduled, featuring research and pedagogy papers, panel discussions, and workshops on teaching economics at the college level (undergraduate and graduate). Registration for the conference and hotel will open on March 1, 2012. For more details to go to ⟨http://www.aeaweb.org/committees/AEACEE/Conference/index.php⟩.

American Economic Association (AEA) 2012 Summer Training Program and Minority Scholarships. After a one-year hiatus, the AEA's Summer Program resumes in 2012, hosted by the Department of Economics and the Robert Wood Johnson Foundation Center for Health Policy at the University of New Mexico. The Program will be held from June 17 through July 28 on the Albuquerque UNM campus. It provides undergraduate students with instruction and research opportunities to enable them to better understand what studying for a Ph.D. in Economics entails. It includes courses in economic theory, mathematics, and econometrics as well as research seminars intended to acquaint students with economic concepts and issues. The Program is open to all qualified students, regardless of race, ethnicity, or gender. Minority fellowships are open to qualified U.S. citizens and permanent residents, with preference for members of historically disadvantaged racial or ethnic minority groups. Additional information, and application and nomination information is at ⟨http://healthpolicy.unm.edu⟩. Send inquiries to ⟨center@unm.edu⟩.

Job Openings for Economists (JOE). The AEA is initiating a listing of retired economists who may be interested in teaching on either a part-time or temporary basis. Individuals can add or delete their name at any time during the period that the listing is open. The listing will be active from February 1 through November 30 each year. Listings will be deleted on November 30; the service will be closed during December and January, re-opening on February 1. Go to ⟨http://www.aeaweb.org/joe/⟩.

Harberger Prize for Retrospective Analysis. *The Journal of Benefit-Cost Analysis (JBCA)* is pleased to announce the Arnold Harberger Prize for Retrospective Analysis. The award and an honorarium of at least $2,000 will go to the best retrospective paper published in the *JBCA*. Papers may be an empirical, retrospective case study in any field and any part of the world, or an advance in methodological thought on retrospective analysis. **Submissions are encouraged by the early summer of 2012.** Please submit though the usual process via the *JBCA* website at ⟨http://www.bepress.com/jbca⟩.

The New York State Economics Association (NYSEA) 65th Annual Conference will be held Friday and Saturday, October 5–6, 2012 at Farmingdale State College. One-page abstracts and proposals for a complete session on a single theme or topic are invited. The **submission deadline is June 15, 2012**. For details go to: ⟨http://nysea.bizland.com/⟩.

Kuhmo Nectar Conference and Summer School on Transportation Economics: Annual conference of the International Transport Economics Association. The Kuhmo Nectar Conference and Summer School 2012 will be held in Berlin, Germany June 18–22, 2012. The aim of the Conference, June 21–22, is to promote scientific excellence in the field of transport economics. Specific topics may include transport investment and funding, congestion pricing, time and risk, agglomeration effects, valuation of intangibles, transport safety, aviation, competition, privatization etc. Prior to the conference, June 18–20, the Kuhmo Nectar Summer School offers PhD students and practitioners a condensed programme introducing recent advances in academic research in transport economics. Go to ⟨www.diw.de/kuhmo2012⟩.

Fellowship in India. The American Institute of Indian Studies invites applications from scholars from all disciplines who wish to conduct their research in India. Junior fellowships are given to doctoral candidates to conduct research for their dissertations in India for up to eleven months. Senior long-term (six to nine months) and short-term (four months or less) fellowships are available for scholars who hold the Ph.D. degree. Scholarly/Professional development fellowships are available to scholars and professionals who have not previously worked in India. Eligible applicants include 1) U.S. citizens; and 2) citizens of other countries who are students or faculty members at U.S. colleges and universities (this rule does not apply to U.S. citizens). Applications can be downloaded from the website: ⟨www.indiastudies.org⟩. Inquiries should be directed to: phone: (773) 702-8638; e-mail: aiis@uchicago.edu. **Application deadline is July 1, 2012.**

The American Economic Association

Correspondence relating to advertising, business matters, permission to quote, or change of address should be sent to the AEA business office: <aeainfo@vanderbilt.edu>. Street address: American Economic Association, 2014 Broadway, Suite 305, Nashville, TN 37203. For membership and subscriptions, go to the AEA website: <http://www.aeaweb.org>. Annual dues for regular membership are $20.00, 30.00, or $40.00, depending on income; for an additional $15.00, you can receive this journal in print. Change of address notice must be received at least six weeks prior to the publication month.

The Journal of
Economic Perspectives

Winter 2012, Volume 26, Number 1

Symposia

Energy Challenges

Higher Education

Articles

AMERICAN ECONOMIC ASSOCIATION