

Sufficient Statistics for Welfare Analysis: A Bridge Between Structural and Reduced-Form Methods

Raj Chetty
Harvard University and NBER

April 2009*

Abstract

The debate between “structural” and “reduced-form” approaches has generated substantial controversy in applied economics. This article reviews a recent literature in public economics that combines the advantages of reduced-form strategies – transparent and credible identification – with an important advantage of structural models – the ability to make predictions about counterfactual outcomes and welfare. This literature has developed formulas for the welfare consequences of various policies that are functions of reduced-form elasticities rather than structural primitives. I present a general framework that shows how many policy questions can be answered by estimating a small set of sufficient statistics using program evaluation methods. I use this framework to synthesize the modern literature on taxation, social insurance, and behavioral welfare economics. Finally, I discuss problems in macroeconomics, labor, development, and industrial organization that could be tackled using the sufficient statistic approach.

*This article has been prepared for the first issue of the *Annual Review of Economics*. E-mail: chetty@fas.harvard.edu. Thanks to Joshua Angrist, David Card, Gary Chamberlain, Liran Einav, Amy Finkelstein, John Friedman, James Hines, Hilary Hoynes, Lawrence Katz, Patrick Kline, Erzo Luttmer, Justin McCrary, Enrico Moretti, Ariel Pakes, Emmanuel Saez, Juan Carlos Suarez Serrato, Glen Weyl, and numerous seminar participants for helpful comments and discussions. Gregory Bruich provided outstanding research assistance. I am grateful for funding from NSF grant SES 0645396.

There are two competing paradigms for policy evaluation and welfare analysis in economics: the “structural” approach and the “reduced-form” approach (also known as the “program evaluation” or “treatment effect” approach). The division between structural and reduced-form methods has split the economics profession into two camps whose research programs have evolved almost independently despite focusing on similar questions. The structural approach specifies complete models of economic behavior and estimates or calibrates the primitives of such models. Armed with the fully estimated model, these studies then simulate the effects of changes in policies and the economic environment on behavior and welfare. This powerful methodology has been applied to an array of topics, ranging from the design of tax and transfer policies in public finance to the sources of inequality in labor economics and antitrust policy in industrial organization.

Critics of the structural approach argue that it is difficult to identify all primitive parameters in an empirically compelling manner because of selection effects, simultaneity bias, and omitted variables. These researchers instead advocate “reduced-form” strategies that estimate statistical relationships, placing priority on identification of causal effects using research designs that exploit quasi-experimental variation.¹ Reduced-form studies have identified a variety of important empirical regularities, especially in labor, public, and development economics. Advocates of the structural paradigm criticize the reduced-form approach for estimating statistics that are not policy invariant parameters of economic models, and therefore have limited relevance for policy and welfare analysis (Rosenzweig and Wolpin 2000, Heckman and Vytlacil 2005, Deaton 2009).²

This paper argues that a set of papers in public economics written over the past decade (see Table 1) provide a middle ground between the two methods. These papers develop “sufficient statistic” formulas that combine the advantages of reduced-form empirics – transparent and credible identification – with an important advantage of structural models – the ability to make precise statements about welfare. The central concept of the sufficient sta-

¹The term “reduced-form” is a misnomer: the relationships that are estimated are generally not reduced-forms of economic models. I use “reduced form” here simply for consistency with standard terminology used to describe design-based studies that identify treatment effects.

²See Section 1 of Rosenzweig and Wolpin (2000) and Table V of Heckman and Vytlacil (2005) for a more detailed comparison of the structural and treatment effect approaches.

tistic approach (illustrated in Figure 1) is to derive formulas for the welfare consequences of policies that are functions of high-level elasticities rather than deep primitives. Even though there are multiple combinations of primitives that are consistent with the inputs to the formulas, all such combinations have the same welfare implications.³ For example, Feldstein (1999) shows that the marginal welfare gain from raising the income tax rate can be expressed purely as a function of the elasticity of taxable income even though taxable income may be a complex function of choices such as hours, training, and effort. Saez (2001) shows that labor supply elasticity estimates can be used to make inferences about the optimal progressive income tax schedule in the Mirrlees (1971) model. Chetty (2008a) shows that the welfare gains from social insurance can be expressed purely in terms of the liquidity and moral hazard effects of the program in a broad class of dynamic, stochastic models. Each of these papers answers a policy question using program evaluation estimates, providing economic meaning to what might otherwise be viewed as “atheoretical” statistical estimates.

The goal of this paper is to elucidate the key concepts of the sufficient statistic methodology and encourage its use as a bridge between structural and reduced-form methods. I first provide a general framework for the derivation of sufficient statistic formulas for welfare analysis. This framework shows how envelope conditions from optimization can be used to reduce the set of parameters that need to be identified. I then illustrate the approach by reviewing several recent papers on tax policy, social insurance, and behavioral public finance.

The idea that it is adequate to estimate certain sufficient statistics rather than the full primitive structure to answer certain questions is not new; it was well understood by the pioneers of structural estimation in the Cowles Commission (Marschak 1953, Koopmans 1953). Indeed, Heckman and Vytlacil (2007) label this idea “Marschak’s maxim,” arguing that “for many decisions (policy problems), only combinations of explicit economic parameters are required – no single economic parameter need be identified.” In early microeconomic work, structural methods were preferred because the parameters of the simple models that

³The term “sufficient statistic” is borrowed from the statistics literature: conditional on the statistics that appear in the formula, other statistics that can be calculated from the same sample provide no additional information about the welfare consequences of the policy.

were being studied could in principle be easily identified.⁴ In the 1980s, it became clear that identification of primitives was difficult once one permitted dynamics, heterogeneity, and selection effects. Concerns about the identification of parameters in these richer models led a large group of empirical researchers to abandon structural methods in favor of more transparent quasi-experimental research designs (e.g., Lalonde 1986, Card 1990, Angrist and Krueger 1991).⁵ A large library of treatment effect estimates was developed in the 1980s and 1990s. The recent sufficient statistic literature maps these treatment effect estimates into statements about welfare in structural models that incorporate dynamics, heterogeneity, and general equilibrium effects.

The structural and sufficient statistic approaches to welfare analysis should be viewed as complements rather than substitutes because each approach has certain advantages. The sufficient statistic method has three benefits. First, it is simpler to implement empirically because less data and variation are needed to identify marginal treatment effects than to fully identify a structural model. Indeed, because of the estimation challenges, structural primitives are often calibrated to match reduced-form moments rather than formally estimated using microdata (Dawkins, Srinivasan, and Whalley 2001). The sufficient statistic approach obviates the need to fully calibrate the structural model. This is especially beneficial in models with heterogeneity and discrete choice, where the set of primitives is very large but the set of marginal treatment effects needed for welfare evaluation remains small. By estimating the relevant marginal treatment effects as a function of the policy instrument, one can integrate the formula for the marginal welfare gain between any two observed values to evaluate policy changes.

Second, identification of structural models often requires strong assumptions given available data and variation. Since it is unnecessary to identify all primitives, sufficient statistic formulas can be implemented under weaker assumptions using design-based empirical methods. The results are therefore more transparent and empirically credible. Third, the

⁴An exception is the work of Harberger (1964), who advocated the use of elasticities as sufficient statistics for tax policy analysis in equilibrium models. As I discuss in section 2, Harberger's work can be viewed as a predecessor to the modern sufficient statistic literature.

⁵Imbens and Wooldridge (2008) review program evaluation methods. Imbens (2009) discusses the advantages of such methods from a statistical perspective.

sufficient statistic approach can be applied even when one is uncertain about the positive model that generates observed behavior – as in recent studies in the behavioral economics literature which document deviations from perfect rationality. In such cases, welfare analysis based on a structural model may be impossible, whereas the more agnostic sufficient statistic approach permits some progress. For instance, Chetty, Looney, and Kroft (2008) derive formulas for the deadweight cost of taxation in terms of price and tax elasticities in a model where agents make arbitrary optimization errors with respect to taxes.

The parsimony of the sufficient statistic approach naturally comes with costs. First, and most important, a new sufficient statistic formula must be derived for each question. For example, Gruber (1997) calculates the optimal *level* of unemployment benefits using a sufficient statistic formula. If one were interested in calculating the optimal *duration* of unemployment benefits, one would have to derive a new formula and estimate a new set of elasticities. In contrast, if one had estimated the structural primitives of the job search model used to derive these formulas, different policy simulations could be conducted with ease. Moreover, for some questions, it may be very difficult to derive a sufficient statistic formula and a structural approach may be the only feasible option.

A second potential weakness of sufficient statistic formulas is that they are more easily misapplied than structural methods. This is because one can draw policy conclusions from a sufficient statistic formula without assessing the validity of the model upon which it is based. For example, Gorodnichenko et al. (2009) show that the assumptions about the costs of evasion underlying Feldstein’s (1999) formula for the excess burden of income taxation are inconsistent with the data. In contrast, because structural methods require full estimation of the model, policy conclusions can only be drawn from models that fit the data.

A common argument in favor of the structural approach is that it has advantages in out-of-sample predictions. One can also make out-of-sample predictions using the sufficient statistic approach by estimating marginal treatment effects as a function of the policy instrument and making a statistical extrapolation. For example, by estimating the elasticity of labor supply with respect to the tax rate at various tax rates, one can extrapolate to the welfare consequences of tax regimes that have not yet been observed. In principle, structural

methods do not require such ad hoc extrapolations, since the primitive structure is by definition policy invariant. However, in practice, structural models often rely on extrapolations based on functional form assumptions (such as constant elasticity utilities). Hence, the real advantage of structural models in out-of-sample predictions comes from the precision of the extrapolation. Statistical extrapolations may be less reliable than extrapolations guided by an economic model that imposes restrictions on how behavior changes with policies.⁶

The structural and sufficient statistic methods can be combined to address the shortcomings of each strategy. For instance, a structural model can be calibrated to match the sufficient statistics that matter for local welfare analysis to improve its empirical credibility. Conversely, when making out-of-sample predictions using a sufficient statistic formula, a structural model can be used to guide the choice of functional forms used to extrapolate the key elasticities. By combining the two methods in this manner, researchers can pick a point in the interior of the continuum between reduced-form and structural estimation, without being pinned to one endpoint or the other. In addition, sufficient statistic formulas provide theoretical guidance for reduced-form empirical work by identifying the parameters of greatest interest for a given question.

The paper is organized as follows. The next section discusses Harberger's (1964) formula for the deadweight cost of taxation, a precursor to the modern literature on sufficient statistics. In Section II, I develop a general framework which provides a six-step rubric for deriving sufficient statistic formulas. Sections III to V review applications of the sufficient statistic method to income taxation, social insurance, and behavioral (non-rational) models. These three sections provide a synthesis of the modern public finance literature, showing how a dozen seemingly unrelated papers are essentially variants on the theme of finding sufficient statistics. The paper concludes in section VI with a discussion of potential applications to fields beyond public finance.

⁶See Lumsdaine, Stock, and Wise (1992) and Keane and Wolpin (1997) for comparisons of reduced-form statistical extrapolations and model-based structural extrapolations. They find that structural predictions are more accurate, but statistical extrapolations that include the key variables suggested by the economic model come quite close.

I A Precedent: Measuring Deadweight Loss

Harberger (1964) popularized the measurement of the excess burden of a commodity tax using a simple elasticity-based formula. This result can be viewed as a precedent to the modern literature on sufficient statistics, and provides a starting point from which to build intuition about the applications discussed below.

Consider a static general equilibrium model in which an individual is endowed with Z units of the numeraire (y), whose price is normalized to 1. Firms convert the numeraire good y (which can be interpreted as labor) into J other consumption goods, $x = (x_1, \dots, x_J)$. Producing x_j units of good j requires an input of $c_j(x_j)$ units of y , where c_j is a weakly convex function. Let $c(x) = \sum_{j=1}^J c_j(x_j)$ denote the total cost of producing a vector x . Production is perfectly competitive. The government levies a unit tax t on good 1. Let $p = (p_1, \dots, p_J)$ the vector of pre-tax prices for the produced goods, which are determined endogenously in market equilibrium.

To simplify the exposition, ignore income effects by assuming that utility is quasilinear in y . The consumer takes the price vector as given and solves:

$$\begin{aligned} \max_{x,y} u(x_1, \dots, x_J) + y \\ \text{s.t. } p \cdot x + tx_1 + y = Z \end{aligned} \tag{1}$$

where $u(x)$ is strictly quasiconcave. The representative firm takes prices as given and solves

$$\max_x p \cdot x - c(x) \tag{2}$$

These two problems define maps from the price vector p to demand and supply of the J goods, $x^D(p)$ and $x^S(p)$. The model is closed by the market clearing condition $x^D(p) = x^S(p)$. Let $p(t)$ denote the market-clearing price vector as a function of the tax rate t .

Suppose the policy maker wants to measure the efficiency cost of the tax t . The efficiency (or “deadweight”) cost of a tax increase equals the loss in surplus from the transactions that fail to occur because of the tax. To calculate the efficiency cost, the conceptual experiment is to measure the net loss in welfare from raising the tax rate and returning the tax revenue

to the taxpayer through a lump-sum rebate.⁷ With quasi-linear utility, the consumer will always choose to allocate the lump-sum rebate to consumption of the numeraire good y . Social welfare can therefore be written as the sum of the consumer's utility (which is a money metric given quasilinearity), producer profits, and tax revenue:

$$\begin{aligned} W(t) &= \left\{ \max_x u(x) + Z - tx_1 - p(t) \cdot x \right\} + \left\{ \max_x p(t) \cdot x - c(x) \right\} + tx_1 \\ &= \left\{ \max_x u(x) + Z - tx_1 - c(x) \right\} + tx_1 \end{aligned} \quad (3)$$

where the second equation effectively recasts the decentralized equilibrium as a planner's allocation problem. In this expression, the term in curly brackets measures private surplus, while the tx_1 term measures tax revenue. The individual treats tax revenue as fixed when choosing x , failing to internalize the effects of his behavior on the lump-sum transfer he ultimately receives. This assumption, which is standard in efficiency cost calculations, captures the intuition that in an economy populated by a large number of individuals, any one individual has a negligible impact on the government revenue and therefore treats it as fixed.

There are two approaches to estimating the effect of an increase in the tax on social welfare ($\frac{dW}{dt}$). The first is to estimate (or calibrate) a J good demand and supply system to recover the utility function $u(x)$ and cost function $c(x)$. Once u and c are known, one can directly compute $W(t)$, recognizing that the tax t will affect equilibrium prices and quantities in all J markets. Preferences can be recovered using the parametric demand systems proposed, for instance, by Stone (1954) or Deaton and Muellbauer (1980). Alternatively, one can fit a demand system to the data and then integrate to obtain the expenditure function, as in Hausman (1981) or Hausman and Newey (1995). The econometric challenge in implementing any of these structural methods is simultaneity: identifying the slope of the supply and demand curves requires $2J$ instruments.

⁷Formally, excess burden is defined using equivalent or compensating variation measures (see e.g., Auerbach 1985). The social welfare calculation here is equivalent to these measures because we are considering a specification without income effects. With income effects, the Harberger formula discussed below applies with the Hicksian elasticity in place of the Marshallian elasticity.

Harberger (1964) proposed an alternative, simpler solution to the problem.⁸ Differentiating (3) and exploiting the first-order conditions from consumer and firm maximization yields

$$\frac{dW(t)}{dt} = -x_1 + x_1 + t \frac{dx_1}{dt} = t \frac{dx_1(t)}{dt} \quad (4)$$

This formula shows that the effect of the tax on equilibrium quantity in the taxed market ($\frac{dx_1(t)}{dt}$) is a “sufficient statistic” for analyzing the efficiency costs of tax changes. By estimating $\frac{dx_1}{dt}(t)$ for different values of t , one can calculate the welfare consequences of any tax change that lies within the observed support of t by integrating (4): $\Delta W = W(t_2) - W(t_1) = \int_{t_1}^{t_2} t \frac{dx_1}{dt}(t) dt$. The key point is that the full system of supply and demand curves does not have to be identified to calculate ΔW .

Harberger’s formula rests upon two conceptual insights that form the basis for modern sufficient statistic applications. First, the behavioral responses ($\frac{dx}{dt}$) in the curly brackets of (3) can be ignored when calculating $\frac{dW}{dt}$ because of envelope conditions from consumer and firm optimization. Intuitively, social welfare has already been optimized by individuals and firms (subject to constraints imposed by the government). Although the tax induces changes in behavior, these behavioral responses cannot have a first-order effect on private welfare; if they did, consumers or firms would not be optimizing. Second, the changes in equilibrium prices ($\frac{dp}{dt}$) can also be ignored when calculating $\frac{dW}{dt}$. This is because prices cancel out of the expression for social welfare in (3). Changes in prices simply redistribute income from producers to consumers without changing aggregate surplus.

These two observations imply that the loss in social surplus from the tax is determined purely by the difference between the agent’s willingness to pay for good x_1 and the cost of producing good x_1 . The difference can be measured by the area between the supply and demand curves and the initial and post-tax quantities (i.e., the “Harberger triangle”), which is proportional to $\frac{dx_1}{dt}$. The total derivative $\frac{dx_1}{dt} = \frac{\partial x_1}{\partial p_1} \frac{\partial p_1}{\partial t} + \frac{\partial x_1}{\partial p_2} \frac{\partial p_2}{\partial t} + \dots + \frac{\partial x_1}{\partial p_J} \frac{\partial p_J}{\partial t}$ measures the effect of an increase in the tax rate on x_1 , allowing all prices and equilibrium demands to change endogenously. The simplicity of Harberger’s approach stems from estimating $\frac{dx_1}{dt}$ directly rather than estimating its various components, which is effectively what the

⁸Hines (1999) colorfully recounts the intellectual history of the deadweight loss triangle.

structural approach requires.

The tradeoffs between the sufficient statistic and structural approaches are apparent in the debate that followed Harberger’s work. One limitation of (4) is that it does not permit pre-existing distortions in the other markets; otherwise the spillover effects would have first-order effects on welfare. This limitation can be addressed by an extension of the formula that includes cross-price elasticities, as shown in Harberger’s original analysis. Under empirically plausible approximations about the structure of the distortions, the formula can be expressed purely in terms of own-price elasticities (Goulder and Williams 2003). Hence, a different sufficient statistic formula is required to handle cases with pre-existing distortions.⁹

A second limitation of (4) is that it cannot be directly used to evaluate counterfactual policy changes such as the imposition of a large new tax on good x_1 . This limitation can be addressed by estimating $\frac{dx_1}{dt}(t)$ for various values of t and making functional-form assumptions to extrapolate out-of-sample. In practice, the Harberger formula is typically implemented under a linear or log-linear approximation to demand (e.g. $\frac{dx_1}{dt}$ constant) because data limitations preclude estimation of higher-order properties of the demand curve. Simulations of calibrated models suggest that implementations of Harberger’s formula using linear approximations are generally quite accurate (Shoven 1976, Ballard, Shoven, and Whalley 1985). Thus, despite its limitations, the simple Harberger “triangle” formula has become central to applied welfare analysis and has inspired a vast literature estimating tax elasticities.

The benefits of Harberger’s approach are especially evident in modern structural models that permit heterogeneity across individuals and discrete choice. I now incorporate these features into the preceding analysis, following Small and Rosen’s (1981) analysis for the discrete choice case.

Extension 1: Heterogeneity. Now suppose the economy has N individuals with heterogeneous preferences. Let x^i denote individual i ’s vector of demands and $x = \sum_{i=1}^N x^i$ denote

⁹A related concern is that one may inadvertently ignore some pre-existing distortions and apply an inaccurate version of the Harberger formula. Indeed, Goulder and Williams argue that previous applications of the simple formula in (4) to assess the deadweight costs of commodity taxation are biased by an order-of-magnitude because they fail to account for interactions with the labor income tax. This mistake would not have been made in a properly specified structural model.

aggregate demand. Individual i is endowed with Z^i units of the numeraire and has utility

$$u^i(x^i) + y \tag{5}$$

Under a utilitarian criterion, social welfare is given by:

$$W(t) = \left\{ \sum_{i=1}^N \max_{x^i} [u^i(x^i) + Z^i - tx_1^i] - c(x) \right\} + t \sum_{i=1}^N x_1^i \tag{6}$$

The structural approach requires identification of the demand functions and utilities for all i agents. The sufficient statistic approach simplifies the identification problem substantially here. Because there is an envelope condition for x^i for every agent, we can ignore all behavioral responses within the curly brackets when differentiating (6) to obtain

$$\frac{dW(t)}{dt} = - \sum_{i=1}^N x_1^i + \sum_{i=1}^N x_1^i + t \frac{d \sum_{i=1}^N x_1^i}{dt} = t \frac{dx_1(t)}{dt} \tag{7}$$

The effect of a tax increase on *aggregate* demand ($\frac{dx_1}{dt}$) is a sufficient statistic for the marginal excess burden of that tax; there is no need to characterize the underlying heterogeneity in the population to implement (7). Intuitively, even though each individual has a different demand elasticity, what matters for government revenue and aggregate welfare is the *total* change in behavior induced by the tax.¹⁰

An important caveat is that with heterogeneity, $\frac{dx_1}{dt}$ may vary considerably with t , since the individuals at the margin will differ with the tax rate. Hence, it is especially important to distinguish average and marginal treatment effects for welfare analysis by estimating $\frac{dx_1}{dt}(t)$ as a function of t in this case.

Extension 2: Discrete Choice. Now suppose individuals can only choose one of the J products $\{1, \dots, J\}$. These products might represent models of cars, modes of transportation, or neighborhoods. Each product is characterized by a vector of K attributes $x_j = (x_{1j}, \dots, x_{Kj})$

¹⁰Of course, to analyze a policy that has heterogeneous impacts across groups, such as a progressive income tax, one needs group-specific elasticity estimates to calculate $\frac{dW}{dt}$. The key point, however, is that the only heterogeneity that matters is at the level of the policy impact; any additional heterogeneity within groups can be ignored. For instance, heterogeneous labor supply responses *within* an income group need to not be characterized when analyzing optimal progressive income taxation.

observed by the econometrician and an unobservable attribute ζ_j . If agent i chooses product j , his utility is

$$u_{ij} = v_{ij} + \varepsilon_{ij}$$

$$\text{with } v_{ij} = Z^i - p_j + \zeta_j + \phi^i(x_j)$$

where ε_{ij} is a random unobserved taste shock. Let P_{ij} denote the probability that individual i chooses option j , $P_j = \sum_i P_{ij}$ denote total (expected) demand for product j , and $P = (P_1, \dots, P_J)$ the vector of aggregate product demands. Product j is produced by competitive firms using $c_j(P_j)$ units of the numeraire good y . Let $c(P) = \sum_j c_j(P_j)$. This model differs from that above in two respects: (1) utility over the consumption goods is replaced by utility over the product attributes $\phi^i(x_j) + \zeta_j + \varepsilon_{ij}$ and (2) the attributes can only be consumed in discrete bundles.

To build intuition, first consider a case in which individuals make binary decisions about whether to buy a single good x_1 ($J = 1$) and the price of x_1 is fixed at p_1 (constant marginal cost of supply). If an individual does not buy x_1 , he spends his wealth on the numeraire and obtains utility $u_i = Z^i$. If he buys x_1 , his utility is $u_i = Z^i - p_1 + \zeta_1 + \phi^i(x_1) + \varepsilon_{i1}$. Let $V^i = \zeta_1 + \phi^i(x_1) + \varepsilon_{i1}$ denote individual i 's gross valuation of x_1 . Let $F(V^i)$ denote the smooth distribution of valuations in the economy and suppose there is a unit mass of agents. Let EZ denote the average level of wealth in the economy.

To calculate $\frac{dW}{dt}$ in this environment, first note that individuals with V^i above a cutoff \bar{V} will buy x_1 . The model is therefore isomorphic to one in which a representative agent chooses \bar{V} to maximize total private surplus. Social surplus can be written as the sum of private surplus and tax revenue:

$$W(t) = \left\{ EZ + \max_{\bar{V}} \int_{\bar{V}}^{\infty} [V^i - (p_1 + t)] dF(V^i) \right\} + t \int_{\bar{V}}^{\infty} dF(V^i) \quad (8)$$

Differentiating this expression and using the envelope condition for \bar{V} , we see that the ag-

gregate demand response remains a sufficient statistic for the marginal welfare gain:

$$\frac{dW}{dt} = -(1 - F(\bar{V})) + (1 - F(\bar{V})) + t \frac{d \int_{\bar{V}}^{\infty} dF(V^i)}{dt} = t \frac{dx_1}{dt}.$$

Intuitively, even though the individual demand functions are discontinuous, the model can be recast as the smooth choice of the demand threshold \bar{V} by a representative agent. This permits application of the standard envelope condition to \bar{V} in the social welfare function. At the microeconomic level, this envelope condition reflects the fact that the marginal individual who stops buying good x_1 when t is raised loses no utility, since he was indifferent about buying good x_1 to begin with.

The same logic applies in the J good case. To simplify exposition and link the results to standard multinomial logit discrete choice models, assume that ε_{ij} has a type 1 extreme value distribution.¹¹ Then it is well known (see e.g. Train 2003) that the probability that a utility-maximizing individual i chooses product j is

$$P_{ij} = \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \quad (9)$$

and that agent i 's expected utility from a vector of prices $p = (p_1, \dots, p_J)$ is

$$S_i(p_1, \dots, p_J) = E \max(u_{i1}, \dots, u_{iJ}) = \log\left(\sum_j \exp v_{ij}\right).$$

Aggregating over the $i = 1, \dots, N$ consumers, (expected) consumer surplus is

$$S = \sum_i \log\left(\sum_j \exp(v_{ij})\right)$$

Since utility is quasilinear, we can add producer profits to this expression to obtain social welfare:

$$W = \sum_i \log\left(\sum_j \exp(v_{ij})\right) + p \cdot P - c(P) \quad (10)$$

¹¹The results that follow do not rely on the assumption that the ε_{ij} errors have an extreme value distribution (Small and Rosen 1981). The distributional assumption simplifies the algebra by yielding a closed-form solution for total surplus, but the envelope conditions used to derive (11) hold with any distribution.

The classical approach to policy analysis in these models is to estimate the distributions of ϕ_i and ζ_j , and simulate total surplus before and after a policy change (see e.g., Train 2003, p60). Identification of such models is challenging, especially if the econometrician does not observe all product attributes, since ζ_j will be correlated with p_j in equilibrium (Berry 1994; Berry, Levinsohn, and Pakes 1995).

Sufficient statistic approaches offer a means of policy analysis that does not require identification of ϕ_i and ζ_j . For example, suppose the government levies a tax t on good 1, raising its price to $p_1 + t$. The government returns the proceeds to agents through a lump-sum transfer T so that y_i becomes $y_i + T$. As above, agents do not internalize the effects of their behavior on the size of the transfer T . Using the envelope condition for profit maximization,

$$\begin{aligned} \frac{dW(t)}{dt} &= \sum_i \left[-\frac{\exp(v_{i1})}{\sum_j \exp(v_{ij})} - \sum_j \frac{dp_j}{dt} \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \right] + \sum_j \frac{dp_j}{dt} P_j + P_1 + t \frac{dP_1}{dt} \quad (11) \\ &= t \frac{dP_1(t)}{dt} \end{aligned}$$

where the second equality follows from (9). Identification of the welfare loss from taxation of good 1 requires estimation of only the effect of the tax on the aggregate market share ($\frac{dP_1}{dt}$), as in the standard Harberger formula.

Now suppose that an ad-valorem tax τ is levied on all the products except the numeraire good, raising the price of product j to $(1 + \tau)p_j$. Again, tax revenue is returned to agents through a lump sum grant. Following a similar derivation,

$$\frac{dW(\tau)}{d\tau} = \tau \sum_j p_j \frac{dP_j(\tau)}{d\tau} = \tau \frac{dE_P(\tau)}{d\tau}$$

where $E_p = \sum_j p_j P_j$ denotes total pre-tax expenditure in the market for the taxed good. The efficiency cost of a tax on all products depends on the aggregate expenditure elasticity for the taxed market; it does not require estimation of the substitution patterns within that market. A similar derivation can be used to show that the efficiency cost of a tax on a single characteristic – such as gas mileage – can be calculated simply by estimating the effect of the tax on the equilibrium quantity of that characteristic (e.g., gasoline consumption).

Hence, many policy questions of interest can be answered simply by estimating reduced-form aggregate demand responses even in discrete choice models.

The modern sufficient statistic literature builds on Harberger’s idea of only identifying the aspects of the model relevant for the question at hand. Before describing specific applications of this approach, I present a general framework that nests the papers in this literature and provides a “recipe” for developing such formulas.

II General Framework

Abstractly, many government policies amount to levying a tax t to finance a transfer $T(t)$. In the context of redistributive taxation, the transfer is to another agent; in the context of social insurance, it is to another state; and in the context of the excess burden calculations above, the transfer can be thought of as being used to finance a public good. I now present a six step rubric for calculating the welfare gain from raising the tax rate t (and the accompanying transfer $T(t)$) using sufficient statistics.

To simplify exposition, the rubric is formally presented in a static model with a single agent. The same sequence of steps can be applied to obtain formulas for multi-agent problems with heterogeneous preferences and discrete choice if $U(\cdot)$ is viewed as a (smooth) social welfare function aggregating the utilities of all the agents, as in (6) and (10). Similarly, dynamics can be incorporated by integrating the utility function over multiple periods.

Step 1: Specify the structure of the model. Let $x = (x_1, \dots, x_J)$ denote the vector of choices for the representative agent in the private sector. A unit tax t is levied on choice x_1 and the transfer $T(t)$ is paid in units of x_J . Let $\{G_1(x, t, T), \dots, G_M(x, t, T)\}$ denote the $M < J$ constraints faced by the agent, which include budget constraints, restrictions on insurance or borrowing, hours constraints, etc. The agent takes t and T as given and makes his choices by solving:

$$\max U(x) \text{ s.t. } G_1(x, t, T) = 0, \dots, G_M(x, t, T) = 0 \tag{12}$$

The solution to (12) gives social welfare as a function of the policy instrument:

$$W(t) = \max_x U(x) + \sum_{m=1}^M \lambda_m G_m(x, t, T)$$

This specification nests competitive production because any equilibrium allocation can be viewed as the choice of a benevolent planner seeking to maximize total private surplus subject to technological constraints.

For example, in the single agent Harberger model analyzed above,

$$\begin{aligned} U(x) &= u(x_1, \dots, x_{J-1}) + x_J \\ G_1(x, t, T) &= T + Z - t_1 x_1 - c(x_1, \dots, x_{J-1}) - x_J. \end{aligned} \tag{13}$$

Note that a more general specification of preferences and constraints will yield a formula that is more robust but harder to implement empirically.

Step 2: Express $\frac{dW}{dt}$ in terms of multipliers. Using the envelope conditions associated with optimization in the private sector, differentiate W to obtain

$$\frac{dW}{dt} = \sum_{m=1}^M \lambda_m \left\{ \frac{\partial G_m}{\partial T} \frac{dT}{dt} + \frac{\partial G_m}{\partial t} \right\} \tag{14}$$

where λ_m denotes the Lagrange multiplier associated with constraint m in the agent's problem in (12). In this equation, $\frac{dT}{dt}$ is known through the government's budget constraint, and $\frac{\partial G_m}{\partial T}$ and $\frac{\partial G_m}{\partial t}$ can be calculated mechanically. For example, in the Harberger model, $T(t) = tx_1$ and hence $\frac{dT}{dt} = x_1 + t \frac{dx_1}{dt}$. Differentiating (13) yields $\frac{dG_1}{dT} = 1$ and $\frac{dG_1}{dt} = -x_1$. It follows that $\frac{dW}{dt} = \lambda_1 t \frac{dx_1}{dt}$.

The critical unknowns are the λ_m multipliers. In the excess burden application, λ_1 measures the marginal value of relaxing the budget constraint. In a social insurance application, λ_1 could represent the marginal value of relaxing the constraint that limits the extent to which agents can transfer consumption across states. If λ_1 is small, there is little value to social insurance, whereas if it is large, $\frac{dW}{dt}$ could be large.

Step 3: Substitute multipliers by marginal utilities. The λ_m multipliers are

recovered by exploiting restrictions from the agent’s first-order-conditions. Optimization leads agents to equate marginal utilities with linear combinations of the multipliers:

$$u'(x_j) = - \sum_{m=1}^M \lambda_m \frac{\partial G_m}{\partial x_j}$$

Inverting this system of equations generates a map from the multipliers into the marginal utilities. It is helpful to impose the following assumption on the structure of the constraints in order to simplify this mapping.

Assumption 1. The tax t enters all the constraints in the same way as the good on which it is levied (x_1) and the transfer T enters all the constraints in the same way as the good in which it is paid (x_J). Formally, there exist functions $k_t(x, t, T)$, $k_T(x, t, T)$ such that

$$\begin{aligned} \frac{\partial G_m}{\partial t} &= k_t(x, t, T) \frac{\partial G_m}{\partial x_1} \forall m = 1, \dots, M \\ \frac{\partial G_m}{\partial T} &= -k_T(x, t, T) \frac{\partial G_m}{\partial x_J} \forall m = 1, \dots, M \end{aligned}$$

Assumption 1 requires that x_1 and t enter every constraint interchangeably (up to a scale factor k_t).¹² That is, increasing t by \$1 and reducing x_1 by \$ k_t would leave all constraints unaffected. A similar interchangeability condition is required for x_J and T . In models with only one constraint per agent, Assumption 1 is satisfied by definition. In the Harberger model, where the only constraint is the budget constraint, k_t corresponds to the mechanical increase in expenditure caused by a \$1 increase in t ($\x_1) vs. a \$1 increase in x_1 ($\$p_1 + t$). Hence, $k_t = \frac{x_1}{p_1 + t}$ in that model. Since increasing the transfer by \$1 affects the budget constraint in the same way as reducing consumption of x_J by \$1, $k_T = 1$.

Models where the private sector choices are second-best efficient subject to the resource constraints in the economy typically satisfy the conditions in Assumption 1. This is because fungibility of resources ensures that the taxed good and tax rate enter all constraints in the same way (see Chetty (2006a) for details). The sufficient statistic approach can be implemented in models that violate Assumption 1 (see section IV for an example), but the

¹²If the tax t is levied on multiple goods (x_1, \dots, x_t) as in Feldstein (1999), the requirement is that it enters the constraints in the same way as the combination of all the taxed goods, i.e. $\frac{\partial G_i}{\partial t} = \sum_{i=1}^t k_t(x, t, T) \frac{\partial G_j}{\partial x_i}$.

algebra is much simpler when this assumption holds. This is because the conditions in Assumption 1 permit direct substitution into (14) to obtain:

$$\begin{aligned}
\frac{dW}{dt} &= \sum_{m=1}^M \lambda_m \left\{ -k_T \frac{\partial G_m}{\partial x_J} \frac{dT}{dt} + k_t \frac{\partial G_m}{\partial x_1} \right\} \\
&= -k_T \frac{dT}{dt} \sum_{m=1}^M \lambda_m \frac{\partial G_m}{\partial x_J} + k_t \sum_{m=1}^M \lambda_m \frac{\partial G_m}{\partial x_1} \\
\frac{dW}{dt} &= k_T \frac{dT}{dt} u'(x_J(t)) - k_t u'(x_1(t)). \tag{15}
\end{aligned}$$

This expression captures a simple and general intuition: increasing the tax t is equivalent to reducing consumption of x_1 by k_t units, which reduces the agent's utility by $k_t u'(x_1(t))$. The additional transfer that the agent gets from the tax increase is $\frac{dT}{dt} k_T$ units of good x_J , which raises his utility by $k_T \frac{dT}{dt} u'(x_J(t))$. Since k_T , k_t , and $\frac{dT}{dt}$ are known based on the specification of the model, this expression distills local welfare analysis to recovering a pair of marginal utilities.¹³

In models with heterogeneity, the aggregate welfare gain is a function of a pair of average marginal utilities across agents. In dynamic models, the welfare gain is also a function of a pair of average marginal utilities, but with the mean taken over the lifecycle for a given agent. This result is obtained using envelope conditions when differentiating the value function.

Step 4: Recover marginal utilities from observed choices. The final step in obtaining an empirically implementable expression for $\frac{dW}{dt}$ is to back out the two marginal utilities. One way to do this is to make assumptions about the relevant marginal utilities based on surveys or external evidence, such as measures of the value of an additional year of healthy life. This is implicitly the approach taken in the cost-benefit analyses sometimes reported in reduced-form studies (e.g., Levitt 1997).

Sufficient statistic studies recover the marginal utilities from choice data using the structure of the model specified in step 1. There is no canned procedure for this step. The applications below provide several illustrations. The trick that is typically exploited is that the marginal utilities are elements in first-order conditions for various choices. As a result,

¹³In many applications, steps 2 and 3 are consolidated into a single step because the constraints can be substituted directly into the objective function.

they can be backed out from the comparative statics of behavior. For instance, in the single agent Harberger model above, the assumption of no income effects implies $u'(x_J) = 1$. To identify $u'(x_1)$, exploit the first-order condition for x_1 , which is $u'(x_1) = p_1 + t$. Plugging in these expressions and the other parameters above into (15), we obtain (4):

$$\frac{dW(t)}{dt} = 1 \cdot (x_1 + t \frac{dx_1}{dt}) - \frac{x_1}{p_1 + t} \cdot (p_1 + t) = t \frac{dx_1(t)}{dt}.$$

Step 5: Empirical Implementation. Suppose the sufficient statistic formula one derives has the following form:

$$\frac{dW}{dt}(t) = f\left(\frac{\partial x_1}{\partial t}, \frac{\partial x_1}{\partial Z}, t\right). \tag{16}$$

When implementing this expression empirically, two issues should be kept in mind. First, the relevant derivatives may require holding different variables fixed depending upon the application. For example, the Harberger formula in (4) calls for measurement of the total derivative $\frac{dx_1}{dt}$, which incorporates general equilibrium effects and price changes in all markets. In contrast, many reduced-form empirical studies explicitly seek to identify the partial derivative $\frac{\partial x_1}{\partial t}$, holding prices in other markets constant. These studies typically compare the behavior of a small group of individuals “treated” by a tax change with other unaffected individuals. Such studies do not recover the sufficient statistic of interest for the Harberger policy question. The elasticities they estimate are, however, directly relevant for assessing the efficiency cost of a tax increase on a small subset of consumers that does not affect equilibrium prices in other markets. The general lesson is that the experiment used to identify the relevant elasticities must be matched to the policy question that is being asked.¹⁴ In situations where one cannot credibly identify the elasticity called for by the formula, it may be possible to make progress using approximations – e.g., a tax levied on a small market has negligible effects on prices in other markets in equilibrium.

The second issue arises from the fact that policy changes of interest are never infinitesimal.

¹⁴A related problem is that reduced-form studies often estimate the combined effect of multi-dimensional changes, such as welfare reforms that affect both time limits and financial incentives to work. Such estimates cannot be directly plugged into sufficient statistic formulas. In such cases, structure may be needed to obtain elasticities with respect to each dimension of the change.

The ideal way to implement (16) to assess the efficiency costs of a discrete policy change is to estimate the inputs as non-parametric functions of the policy instrument t . With estimates of $\frac{\partial x_1}{\partial t}(t)$ and $\frac{\partial x_1}{\partial Z}(t)$, one can integrate (16) between any two tax rates t_1 and t_2 that lie within the support of observed policies to evaluate the welfare gain ΔW for a policy change of interest. This procedure is similar in spirit to Heckman and Vytlacil's (2001, 2005) recommendation that researchers estimate a complete schedule of marginal treatment effects (MTE), and then integrate that distribution over the desired range to obtain policy relevant treatment effects. In the present case, the marginal welfare gain at t depends on the MTE at t ; analysis of non-marginal changes requires estimation of the MTE as a function of t . Predictions about welfare gains from policies outside the observed support can be made by extrapolating $\frac{\partial x_1}{\partial t}(t)$ out of sample. That is, one can effectively obtain sufficient statistics for out-of-sample welfare analysis under assumptions about the functional form of $\frac{\partial x_1}{\partial t}(t)$.

In most applications, there is insufficient power to estimate $x_1(t)$ non-parametrically. Instead, typical reduced-form studies estimate local average treatment effects (Angrist, Graddy, and Imbens 2000), such as the effect of a discrete change in the tax rate from t_1 to t_2 on demand: $\frac{\Delta x_1}{\Delta t} = \frac{x_1(t_2) - x_1(t_1)}{t_2 - t_1}$. The estimate of $\frac{\Delta x_1}{\Delta t}$ permits inference about the effect of raising the tax rate from t_1 to t_2 on welfare. To see this, consider the Harberger model, where $\frac{dW}{dt}(t) = t \frac{dx_1}{dt}(t)$. A researcher who has estimated $\frac{\Delta x_1}{\Delta t}$ has two options. This first is to bound the average welfare gain over the observed range:

$$\begin{aligned} W(t_2) - W(t_1) &= \int_{t_1}^{t_2} \frac{dW}{dt} dt = \int_{t_1}^{t_2} t \frac{dx_1}{dt}(t) dt \\ \Rightarrow t_1 \frac{\Delta x_1}{\Delta t} &> \overline{dW/dt} > t_2 \frac{\Delta x_1}{\Delta t} \end{aligned} \tag{17}$$

Intuitively, the excess burden of taxation depends on the slope of the $x_1(t)$ between t_1 and t_2 , multiplied by the height of the ‘‘Harberger trapezoid’’ at each point. When one observes only the average slope between the two tax rates, bounds on excess burden can be obtained by setting the height to the lowest and highest points over the interval.

The second option is to use an approximation to $x_1(t)$ to calculate $\overline{dW/dt}$. For instance, if one can estimate only the first-order properties of $x_1(t)$, making the approximation that $\frac{dx_1}{dt}$

is constant over the observed range implies

$$\overline{dW/dt} \simeq \frac{t_1 + t_2}{2} \frac{\Delta x_1}{\Delta t}$$

If $x_1(t)$ is linear, the average height of the trapezoid and $\frac{\Delta x_1}{\Delta t}$ exactly determine excess burden. If one has adequate data and variation to estimate higher-order terms of $x_1(t)$, these estimates can be used to fit a higher order approximation to $x_1(t)$ to obtain a more accurate estimate of $\overline{dW/dt}$.

The same two options are available in models in which $\frac{dW}{dt}$ is a function of more than one behavioral response, as in (16). Bounds may be obtained using the estimated treatment effects $(\frac{\Delta x_1}{\Delta Z}, \frac{\Delta x_1}{\Delta t})$ by integrating $\frac{dW}{dt}$ and setting the other parameters at their extrema as in (17). Under a linear approximation to demand ($\frac{dx_1}{dt}, \frac{dx_1}{dZ}$ constant), treatment effects can be mapped directly into the marginal welfare gain: $\frac{dW(t)}{dt} = f(\frac{\Delta x_1}{\Delta t}, \frac{\Delta x_1}{\Delta Z}, t)$. If $\frac{dW(t)}{dt}$ can only be estimated accurately at the current level of t , one can at least determine the direction in which the policy instrument should be shifted to improve welfare.

The bottom line is that the precision of a sufficient statistic formula is determined by the precision of the information available about the sufficient statistics as a function of the policy instrument. In the applications discussed below, the data and variation available only permit estimation of first-order properties of the inputs, and the authors are therefore constrained to calculating a first-order approximation of $\overline{dW/dt}$. The potential error in this linear approximation can be assessed using the bounds proposed above or using a structural model.

Step 6. Model Evaluation. Although sufficient statistic formulas do not require full specification of the model, they do require some modelling assumptions; it is impossible to make theory-free statements about welfare. It is important to assess the validity of these assumptions to ensure that the formula's results are accurate. Unfortunately, because this step is not necessary to calculate $\frac{dW(t)}{dt}$, it is often not implemented in existing sufficient statistic studies.

The model can be evaluated in two ways. First, one can test qualitative predictions that would falsify the assumptions that are central for deriving the sufficient statistic formula.

For example, Harberger’s formula assumes that individuals treat prices and taxes identically, making choices based only on the total price of the good ($p + t$). This assumption can be tested by comparing price and tax elasticities of demand (Chetty, Looney, and Kroft 2008). Second, one should identify at least one vector of structural parameters ω that is consistent with the sufficient statistics estimated in step 5. If the empirical estimates of the sufficient statistics are internally consistent with the model, at least one ω must fit the estimated statistics. This is not a stringent test, because the idea of the sufficient statistic approach is that there will be multiple values of ω consistent with the sufficient statistics. However, there are cases in which the estimated high-level elasticities may not be consistent with any underlying vector of primitives, rejecting the assumptions of the model (see Chetty (2006a,b) for an example).

The next three sections show how a variety of recent papers in public economics can be interpreted as applications of this six-step rubric. Each application illustrates different strengths and weaknesses of the sufficient statistic approach and demonstrates the techniques that are helpful in deriving such formulas.

III Application 1: Income Taxation

Since the seminal work of Mirrlees (1971) and others, there has been a large structural literature investigating the optimal design of income tax and transfer systems. Several studies have simulated optimal tax rates in calibrated versions of the Mirrlees model (see Tuomala 1990 for a survey). A related literature uses microsimulation methods to calculate the effects of changes in transfer policies on behavior and welfare. The most recent structural work in this area has generalized the Mirrlees model to dynamic settings and simulated the optimal design of tax policies in such environments using calibrated models. Parallel to this literature, a large body of work in labor economics has investigated the effects of tax and transfer programs on behavior using program evaluation methods. See Table 1 for examples of structural and reduced-form studies.

Recent work in public economics has shown that the elasticities estimated by labor economists can be mapped into statements about optimal tax policy in the models that have been

analyzed using structural methods. This sufficient statistic method has been widely applied in the context of income taxation in the past decade, with contributions by Feldstein (1995, 1999), Piketty (1997), Diamond (1998), Saez (2001), Gruber and Saez (2002), Goulder and Williams (2003), Chetty (2008b), and others. All of these papers can be embedded in the general framework proposed above. I focus on two papers here in the interest of space: Feldstein (1999) and Saez (2001).

Feldstein (1999). Traditional empirical work on labor supply did not incorporate the potential effects of taxes on choices other than hours of work. For instance, income taxes could affect an individual's choice of training, effort, or occupational choice. Moreover, individuals may be induced to shelter income from taxation by evading or avoiding tax payments (e.g. taking fringe benefits, underreporting earnings). While some studies have attempted to directly examine the effects of taxes on each of these margins, it is difficult to account for all potential behavioral responses to taxation by measuring each channel separately.

Feldstein proposes an elegant solution to the problem of calculating the efficiency costs of taxation in a model with multi-dimensional labor supply choices. His insight is that the elasticity of taxable income with respect to the tax rate is a sufficient statistic for calculating deadweight loss. Feldstein considers a model in which an individual makes J labor supply choices (x_1, \dots, x_J) that generate earnings. Let w_j denote the wage paid for choice j and $\psi_j(x_j)$ denote the disutility of labor supply through margin x_j .¹⁵ In addition, suppose that the agent can shelter $\$e$ of earnings from the tax authority (via sheltering or evasion) by paying a cost $g(e)$. Total taxable income is $TI = \sum_{j=1}^J w_j x_j - e$. Let $c = (1 - t)TI + e$ denote consumption. For simplicity, assume that utility is linear in c to abstract from income effects. Feldstein shows that it is straightforward to allow for income effects. As in the Harberger model, we calculate the excess burden of the tax by assuming that the government returns the tax revenue to the agent as a lump sum transfer $T(t)$. Using the

¹⁵Feldstein takes the wage rates w_j as fixed, implicitly assuming that the demand for each type of labor supply is perfectly elastic. Allowing for downward-sloping labor demand curves in a competitive market does not affect the formula he derives for the same reason that endogenous prices do not affect the Harberger formula in (4). With endogenous wages, the sufficient statistic for deadweight loss is the total derivative $\frac{dT I}{dt}$, which incorporates all equilibrium wage responses to the tax change.

notation introduced in section II, we can write this model formally as:

$$\begin{aligned} u(c, x, e) &= c - g(e) - \sum_{j=1}^J \psi_j(x_j) \\ T(t) &= t \cdot TI \\ G_1(c, x, t) &= T + (1 - t)TI + e - c \end{aligned}$$

Social welfare is

$$W(t) = \left\{ (1 - t)TI + e - g(e) - \sum_{j=1}^J \psi_j(x_j) \right\} + t \cdot TI \quad (18)$$

To calculate the marginal excess burden $\frac{dW}{dt}$, totally differentiate (18) to obtain

$$\begin{aligned} \frac{dW}{dt} &= TI + t \frac{dTI}{dt} - TI + (1 - t) \frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^J \psi'_j(x_j) \frac{dx_j}{dt} \\ &= \frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^J \psi'_j(x_j) \frac{dx_j}{dt} \end{aligned} \quad (19)$$

This equation is an example of the marginal utility representation in (15) given in step 3 of the rubric in section II. To recover the marginal utilities (step 4), Feldstein exploits the first order conditions

$$\begin{aligned} g'(e) &= t \\ \psi'_j(x_j) &= (1 - t)w_j \\ \Rightarrow \sum_{j=1}^J \psi'_j(x_j) \frac{dx_j}{dt} &= \sum_{j=1}^J (1 - t)w_j \frac{dx_j}{dt} = (1 - t) \frac{d(TI + e)}{dt} \end{aligned} \quad (20)$$

where the last equality follows from the definition of TI . Plugging these expressions into (19) and collecting terms yields the following expression for the marginal welfare gain from raising the tax rate from an initial rate of t :

$$\frac{dW(t)}{dt} = t \frac{dTI(t)}{dt}. \quad (21)$$

A simpler, but less instructive, derivation of (21) is to differentiate (18), recognizing that behavioral responses have no first-order effect on private surplus (the term in curly brackets) because of the envelope conditions. This immediately yields $\frac{dW}{dt} = -TI + TI + t\frac{dTI}{dt}$.

Equation (21) shows that we simply need to measure how taxable income responds to changes in the tax rate to calculate the deadweight cost of income taxation. There is no need to determine whether TI changes because of hours responses, changes in occupation, or avoidance behaviors in order to calculate efficiency costs. Intuitively, the agent supplies labor on every margin (x_1, \dots, x_J) up to the point where his marginal disutility of earning another dollar through that margin equals $1 - t$. The marginal social value of earning an extra dollar net of the disutility of labor is therefore t for all margins. Likewise, the agent optimally sets the marginal cost of reporting \$1 less to the tax authority ($g'(e)$) equal to the marginal private value of doing so (t). Hence, the marginal social costs of reducing earnings (via any margin) and reporting less income via avoidance are the same at the individual's optimal allocation. This makes it irrelevant which mechanism underlies the change in TI for efficiency purposes.

The main advantage of identifying $\frac{dTI(t)}{dt}$ as a sufficient statistic is that it permits inference about efficiency costs without requiring identification of the potentially complex effects of taxes on numerous labor supply, evasion, and avoidance behaviors. Moreover, data on taxable income are available on tax records, facilitating estimation of the key parameter $\frac{dTI}{dt}$. Feldstein (1995) implements (21) by estimating the changes in reported taxable income around the Tax Reform Act of 1986, implicitly using the linear approximation described in step 5 of the rubric. He concludes based on these estimates that the excess burden of taxing high income individuals is very large, possibly as large as \$2 per \$1 of revenue raised. This result has been influential in policy discussions by suggesting that top income tax rates should be lowered (see e.g., Joint Economic Committee 2001). Subsequent empirical work motivated by Feldstein's result has found smaller values of $\frac{dTI}{dt}$, and the academic debate about the value of the taxable income elasticity remains active.

The sixth step of the rubric – model evaluation – has only been partially implemented in the context of Feldstein's formula. Slemrod (1995) and several other authors have found that the large estimates of $\frac{dTI}{dt}$ are driven primarily by evasion and avoidance behaviors ($\frac{de}{dt}$).

However, these structural parameters $(g(e), \psi_j(x_j))$ of the model have not been directly evaluated. Chetty (2008b) gives an example of the danger in not investigating the structural parameters. Chetty argues that the marginal social cost of tax avoidance may not be equal to the tax rate at the optimum – violating the first-order-condition (20) that is critical to derive (21) – for two reasons. First, some of the costs of evasion and avoidance constitute transfers, such as the payment of fines for tax evasion, rather than resource costs. Second, there is considerable evidence that individuals overestimate the true penalties for evasion. Using a sufficient statistic approach analogous to that above, Chetty relaxes the $g'(e) = t$ restriction and obtains the following generalization of Feldstein’s formula:

$$\frac{dW(t)}{dt} = t\left\{\mu(t)\frac{dT I(t)}{dt} + (1 - \mu(t))\frac{dL I(t)}{dt}\right\} \quad (22)$$

where $L I = \sum_{j=1}^J w_j x_j$ represents total earned income and $\mu(t) = \frac{g'(e(t))}{t}$ measures the gap between social marginal costs of avoidance and the tax rate. Intuitively, deadweight loss is a weighted average of the taxable income elasticity $(\frac{dT I}{dt})$ and the total earned income elasticity $(\frac{dL I}{dt})$, with the weight determined by the resource cost of sheltering. If avoidance does not have a large resource cost, changes in e have little efficiency cost, and thus it is only $\frac{dL I}{dt}$ – the “real” labor supply response – that matters for deadweight loss.

Not surprisingly, implementing Chetty’s more general formula requires identification of more parameters than Feldstein’s formula. The most difficult parameter to identify is $g'(e)$, which is a marginal utility. By leaving $g'(e)$ in the formula, Chetty does not complete step 4 of the rubric above; as a result, further work is required to implement (22). Gorodnichenko et al. (2009) provide a method of recovering $g'(e)$ from consumption behavior. Their insight is that real resource costs expended on evasion should be evident in consumption data; thus, the gap between income and consumption measures can be used to infer $g'(e)$. Implementing this method to analyze the efficiency costs of a large reduction in income tax rates in Russia, Gorodnichenko et al. find that $g'(e)$ is quite small and that $\frac{dT I}{dt}$ is substantial, whereas $\frac{dL I}{dt}$ is not. They show that Feldstein’s formula substantially overestimates the efficiency costs of taxation relative to Chetty’s more general measure. Intuitively, reported taxable incomes are highly sensitive to tax rates, but the sensitivity is driven by avoidance behavior that has

little social cost at the margin and hence does not reduce the total size of the pie significantly.

This literature underscores the point that sufficient statistic approaches are not model free. It is critical to evaluate the structure of the model, even though the formula for $\frac{dW}{dt}$ can be implemented without the last step of the rubric. In the taxable income application, estimating $g'(e)$ has value instead of simply assuming that $g'(e) = t$ given plausible concerns that this condition does not hold in practice.

Saez (2001). Harberger and Feldstein study the efficiency effects and optimal design of a linear tax. Much of the literature on optimal income taxation has focused on non-linear income tax models and the optimal progressivity of such systems. Mirrlees (1971) formalizes this question as a mechanism design problem, and provides a solution in differential equations that are functions of primitive parameters. The Mirrlees solution offers little intuition into the forces that determine optimal tax rates. Building on the work of Diamond (1998), Saez (2001) expresses the optimality conditions in the Mirrlees model in terms of empirically estimable sufficient statistics.

Saez analyzes a model in which individuals choose hours of work, l , and have heterogeneous wage rates w distributed according to a distribution $F(w)$. Wage rates (skills) are unobservable to the government. Let pre-tax earnings be denoted by $z = wl$. For simplicity, I again restrict attention to the case without income effects, as in Diamond (1998).

Saez begins by analyzing the optimal tax rate on top incomes. He considers a model where the government levies a linear tax τ on earnings above a threshold \bar{z} and characterizes the properties of the optimal tax rate τ^* as $\bar{z} \rightarrow \infty$. For a given \bar{z} , individuals maximize utility

$$\begin{aligned} u(c, l) &= c - \psi(l) \\ \text{s.t. } G_1(c, l) &= (1 - \tau) \max(wl - \bar{z}, 0) + \bar{z} - c = 0 \end{aligned}$$

Let $c(w, \tau)$ and $l(w, \tau)$ denote an agent's optimal choices as a function of his wage and the tax rate and $z(w, \tau) = wl(w, \tau)$ denote the optimized earnings function. Let $z_m(\bar{z}) = E[wl(w, \tau) | z(w, \tau) > \bar{z}]$ denote the mean level of earnings for individuals in the top bracket. Let \bar{w} denote the wage threshold that corresponds to an earnings threshold of \bar{z} when the tax

rate is τ : $\bar{w}l(\bar{w}, \tau) = \bar{z}$. The tax revenue generated by the top bracket tax is $R = \tau(z_m(\bar{z}) - \bar{z})$. The planner uses this tax revenue to fund a project that has a (normalized) value of \$1 per dollar spent.

The social planner's objective is to maximize a weighted average of individual's utilities, where the weights $\tilde{G}(u)$ are "social welfare weights" that reflect the redistributive preferences of the planner:

$$W = \left\{ \int_0^{\infty} \tilde{G}(u(c(w, \tau), wl(w, \tau))) dF(w) \right\} + \tau(z_m(\bar{z}) - \bar{z})$$

In this equation, the first term (in curly brackets) represents private surplus and the second term reflects government revenue. To calculate $\frac{dW}{d\tau}$, observe that individuals with incomes below \bar{z} are unaffected by the tax increase. Normalize the measure of individuals in the top bracket to 1. Utility maximization implies that behavioral responses ($\frac{\partial l}{\partial \tau}$) have no first-order effect on private surplus, as $wu_c(w, \tau) = \psi'(l(w, \tau))$. Using this envelope condition, we obtain

$$\begin{aligned} \frac{dW}{d\tau}(\tau) &= -\int_{\bar{w}}^{\infty} \tilde{G}_u(u)(z(w, \tau) - \bar{z}) dF(w) + [(z_m - \bar{z}) + \tau \frac{dz_m}{d\tau}] \\ &= -(z_m(\bar{z}) - \bar{z})\bar{g} + [(z_m(\bar{z}) - \bar{z}) + \tau \frac{dz_m}{d\tau}] \end{aligned} \quad (23)$$

where $\bar{g} = \frac{\int_{\bar{w}}^{\infty} \tilde{G}_u(u)(z - \bar{z}) dF(w)}{\int_{\bar{w}}^{\infty} (z - \bar{z}) dF(w)}$ denotes the mean marginal social welfare weight placed by the planner on individuals in the top tax bracket. The parameter \bar{g} measures the social value of giving \$1 more income to individuals in the top bracket relative to the value of public expenditure. If $\bar{g} = 1$, the government weighs the consumption of the individuals it taxes and public expenditure equally, and (23) collapses to the Harberger formula for excess burden in (4). When $\bar{g} < 1$, the first term in (23) captures the welfare loss to individuals in the top tax bracket from having to pay more taxes. The second term reflects the gain in revenue to the government, which consists of two familiar components: the mechanical gain in revenue and the offset due to the behavioral response.

Equation (23) shows that three parameters are together sufficient statistics for the welfare gain of increasing top income tax rates: (1) the effect of tax rates on earnings ($\frac{dz_m}{d\tau}$),

which quantifies the distortions created by the tax; (2) the shape of the earnings distribution ($z_m(\bar{z})$), which measures the mass of individuals whose behavior is distorted by the marginal tax, and (3) the marginal social welfare weight (\bar{g}), which measures the planner's redistributive preferences. Note that Saez does not implement step 4 of the rubric – recovery of marginal utilities from observed behavior – because he views the relevant marginal utility in this case (\bar{g}) as a feature of the planner's social welfare function that is external to the choice environment. Thus, \bar{g} is determined by the shape of the earnings distribution and the (exogenous) specification of the social welfare function (e.g. utilitarian or Rawlsian).

The advantage of (23) relative to a structural approach is that one does not need to identify preferences (ψ) or the shape of the skill distribution $F(w)$ to calculate $dW/d\tau$. Moreover, one can permit arbitrary heterogeneity across skill types in preferences without changing the formula. The disadvantage of (23) is that z_m, \bar{g} , and $\frac{dz_m}{d\tau}$ are endogenous to τ : the level of earnings and the weight the social planner places on top earners presumably decrease with τ , while $\frac{dz_m}{d\tau}$ may vary with τ depending upon the shape of the $\psi(l)$ function. Hence, $\frac{dW}{d\tau}(\tau)$ measures only the marginal welfare gain at a given tax rate τ and must be estimated at all values of τ to calculate the tax rate τ^* that maximizes W . To simplify empirical implementation and derive an explicit formula for the optimal tax rate, Saez observes that the ratio $\frac{z_m(\bar{z})}{\bar{z}}$ is approximately constant in the upper tail of the empirical distribution of earnings in the U.S.: that is, the upper tail of the income distribution is well described by a Pareto distribution. A Pareto distribution with parameter a has $\frac{z_m(\bar{z})}{\bar{z}} = \frac{a}{a-1}$ for all \bar{z} . Hence, (23) can be expressed as

$$\frac{dW}{d\tau} = \frac{(1 - \bar{g})}{a - 1} \bar{z} + \tau \frac{dz_m}{d\tau}$$

The optimal top-bracket tax rate τ satisfies $\frac{dW}{d\tau}(\tau) = 0$, implying

$$\frac{\tau^*}{1 - \tau^*} = \frac{1 - \bar{g}}{a\varepsilon} \tag{24}$$

where $\varepsilon = \frac{dz_m}{d(1-\tau)} \frac{1-\tau}{z_m}$ denotes the taxable income elasticity in the top bracket. In the Mirrlees model, a and ε converge to constants (invariant to τ) in the limit as $\bar{z} \rightarrow \infty$. Equation (24) is therefore an *explicit* formula for the optimal asymptotic top income tax rate if the

social welfare weight \bar{g} is taken as exogenous. For example, one plausible assumption is that $\bar{g} \rightarrow 0$ as $\bar{z} \rightarrow \infty$. Saez exploits this property of (24) to calculate optimal top income tax rates using reduced-form estimates of the taxable income elasticity for high incomes (Gruber and Saez 2002) and a Pareto parameter of $a = 2$ consistent with the earnings distribution in the U.S. He finds that optimal top income tax rates are generally above 50% when the formula is calibrated using plausible elasticities.

Building on this sufficient statistic approach, Saez characterizes the optimal tax rate at any income level z in a non-linear tax system. Let E denote the fixed amount of government expenditure that must be financed through taxation. Let $T(z)$ denote the total tax paid by an individual who earns income z , so that net of tax income is $z - T(z)$. Let $\varepsilon(z) = \frac{dz}{d(1-\tau)} \frac{1-\tau}{z}$ denote the earnings elasticity at income level z and $h(z)$ the density of the earnings distribution at z . Finally, let $\tilde{G}(u(z))$ denote the weight that the planner places on an individual with earnings z and $g(z) = \tilde{G}_u \cdot u_c(z)$ the corresponding marginal social welfare weight.

The government chooses the schedule $T(Z)$ that maximizes social welfare

$$W(T(z)) = \int_0^{\infty} \tilde{G}(u(c(w, T), wl(w, T))) dF(w)$$

subject to resource and incentive-compatibility constraints:

$$\begin{aligned} G_1(c, z, T) &= \int_0^{\infty} z(w, T) dF(w) - \int_0^{\infty} c(w, T) dF(w) - E = 0 \\ G_2(c, z, T) &= (1 - T'(z))w - \psi'(l(w)) = 0 \end{aligned}$$

Exploiting envelope conditions and perturbation arguments as above, the first order conditions for the optimal tax rates can be expressed in terms of sufficient statistics. In the case without income effects, the optimal tax schedule satisfies the following condition at all z :

$$\frac{T(z)}{1 - T(z)} = \frac{1}{\varepsilon(z)zh(z)} \int_z^{\infty} (1 - g(z'))h(z')dz' \quad (25)$$

Equation (25) depends on the same three parameters as (23): the taxable income elasticity, the shape of the earnings distribution, and the social welfare weights. It is again important

to recognize that all three of these parameters are endogenous to the tax regime itself, and hence (25) is not an explicit formula for optimal taxation. The techniques used to obtain the explicit formula for the asymptotic top income tax rate in (24) cannot be applied at an arbitrary income level z because there are no analogous limit convergence results. Hence, (25) can only be used to evaluate how perturbing the existing tax system $T(z)$ would affect welfare.

To go further and calculate the optimal tax schedule, Saez makes additional assumptions about the model's structure. He assumes that the elasticity $\varepsilon(z)$ is constant, which pins down the functional form of the utility $u(c, l)$. Given this utility function and an elasticity estimate taken from the literature, he infers the skill distribution $F(w)$ from the empirically observed distribution of incomes in the current tax regime. Having identified all the primitives of the model, he simulates the optimal tax schedule in the calibrated model. The resulting optimal income tax schedule is inverse-U shaped, with a large lump sum grant to non-workers and marginal rates ranging from 50-80%.

This exercise illustrates the costs and benefits of placing more structure on the problem. By assuming a one-parameter constant-elasticity utility, one can point identify the primitives from the data and calculate the optimal tax schedule. However, the strong assumptions required for structural identification of the model reduce our confidence in the optimal tax schedule calculations. We have greater confidence in the top tax rate calculations based on the sufficient statistic formula in (24) or marginal welfare gain calculations based on (25).

IV Application 2: Social Insurance

Programs such as unemployment insurance, health insurance, social security, workers compensation, and disability insurance account for the majority of government expenditure in many countries. Starting with seminal contribution of Wolpin (1987), a large literature has studied the optimal design of social insurance programs in dynamic structural models. Parallel to this literature, a large body of reduced-form empirical work has investigated the impacts of social insurance programs on health expenditures, unemployment durations, consumption, disability claims, etc. See Table 1 for examples of structural and reduced-form

studies.

In the context of social insurance, an important harbinger to the sufficient statistic approach is the work of Baily (1978), who showed that the optimal level of unemployment benefits can be expressed as a function of a small set of parameters in a static model. Baily’s result was viewed as being of limited practical relevance because of the strong assumptions made in deriving the formula. However, recent work has shown that the parameters Baily identified are actually sufficient statistics for welfare analysis of social insurance in a rich class of dynamic models. Studies in this literature include Gruber (1997), Chetty (2006a), Shimer and Werning (2007), Chetty (2008a), Einav, Finkelstein, and Cullen (2008), and Chetty and Saez (2008). I now embed these papers in the general framework above, focusing primarily on the first four papers.¹⁶

Baily (1978) and Chetty (2006a). For simplicity, consider a static model with two states: high and low. Let w_h denote the individual’s income in the high state and $w_l < w_h$ income in the low state. Let A denote wealth. Let c_h denote consumption in the high state and c_l consumption in the low state. The low state can be thought of as corresponding to job loss, injury, disability, natural disaster, etc. The agent can control the probability of being in the bad state by exerting effort e at a cost $\psi(e)$. For instance, “effort” could reflect searching for a job, taking precautions to avoid injury, or locating a house away from areas prone to natural disasters. Choose units of e so that the probability of being in the high state is given by $p(e) = e$.

Individuals may have some ability to insure against shocks through informal private sector arrangements, such as transfers between relatives. To model such informal private-insurance arrangements, suppose that the agent can transfer $\$b_p$ between states at a cost $q(b_p)$, so that increasing consumption by b_p in the low state requires payment of a premium $\frac{1-e}{e}b_p + q(b_p)$ in the high state. The loading factor $q(b_p)$ can be interpreted as the degree of incompleteness in private insurance. If $q(b_p) = 0$, private insurance markets are complete; if $q(b_p) = \infty$, there is no capacity for private insurance.

¹⁶The tax and social insurance problems are closely related because social insurance is effectively state-contingent taxation. Rather than levying taxes on the basis of income, taxes and transfers are levied on the basis of a state (joblessness, sickness, injury, disability, etc.). Conversely, redistributive taxation is social insurance against uncertain skill realizations behind the veil of ignorance.

The government pays a benefit b in the low state that is financed by an actuarially fair tax $t(b) = \frac{1-e}{e}b$ in the high state. The model can be formally specified using the notation in section II as follows.¹⁷

$$\begin{aligned}
U(c_l, c_h, e) &= eu(c_h) + (1-e)u(c_l) - \psi(e) \\
t(b) &= \frac{1-e}{e}b \\
G_1(c_l, c_h, t) &= c_h + \frac{1-e}{e}b_p + q(b_p) + t - w_h - A \\
G_2(c_l, c_h, t) &= c_l - b_p - b - w_l - A
\end{aligned} \tag{26}$$

Substituting the constraints into the utility function yields social welfare as a function of the government benefit level:

$$W(b) = eu(A + w_h - \frac{1-e}{e}b_p - q(b_p) - t(b)) + (1-e)u(A + w_l + b_p + b) - \psi(e)$$

Differentiating this expression and using the envelope conditions for b_p and e gives

$$\begin{aligned}
\frac{dW(b)}{db} &= (1-e)u'(c_l) - \frac{dt}{db}eu'(c_h) \\
&= (1-e)\left\{u'(c_l) - \left(1 + \frac{\varepsilon_{1-e,b}}{e}\right)u'(c_h)\right\}
\end{aligned}$$

where $\varepsilon_{1-e,b} = \frac{d(1-e)}{db} \frac{b}{1-e}$ denotes the elasticity of the probability of being in the bad state (which can be measured as the unemployment rate, rate of health insurance claims, etc.) with respect to the benefit level. This elasticity measures the total effect of an increase in benefits on e , taking into account the tax increase needed to finance the higher level of benefits.

In tax models with quasilinear utility, the welfare gain measure $\frac{dW}{dt}$ is a money metric. Since curvature of utility is an essential feature of the social insurance problem, we need a method of converting $\frac{dW}{db}$ to a money metric. An intuitive metric is to normalize the welfare gain from a \$1 (balanced budget) increase in the size of the government insurance program

¹⁷I follow the convention in the social insurance literature of specifying the problem in terms of the transfer benefit b rather than the tax t .

by the welfare gain from raising the wage bill in the high state by \$1:

$$\begin{aligned}
 M_W(b) &= \frac{\frac{dW}{db}(b)/(1-e)}{\frac{dW}{dw_h}(b)/e} \\
 &= \frac{u'(c_l) - u'(c_h)}{u'(c_h)} - \frac{\varepsilon_{1-e,b}}{e}
 \end{aligned} \tag{27}$$

This expression, which is Baily’s (1978) formula, corresponds to the marginal utility expression obtained after the third step of the rubric in section II. The first term in (27) measures the gap in marginal utilities between the high and low states, which quantifies the welfare gain from transferring an additional dollar from the high to low state. The second term measures the cost of transferring this \$1 due to behavioral responses.

Chetty (2006a) establishes that the parameters in (27) are “sufficient statistics” in that they are adequate to calculate $M_W(b)$ in a general class of dynamic models that nest existing structural models of insurance. Chetty analyzes a dynamic model where transitions from the good state to the bad state follow an arbitrary stochastic process. Agents make J choices and are subject to M constraints. The choices could include variables such as reservation wages, savings behavior, labor supply, or human capital investments. Subject to a regularity condition analogous to Assumption 1, Chetty shows that (27) holds in this general model, with the difference in marginal utilities replaced by the difference between the average marginal utilities in the high and low states over the agent’s life. This result distills the calculation of welfare gains in complex dynamic models to two parameters: the gap in average marginal utilities and the elasticity that enters the government’s budget constraint $\varepsilon_{1-e,b}$. Identification of parameters such as asset limits or the degree of private insurance ($q(b_p)$) is not required. This permits calculation of $\frac{dW}{db}$ without the assumptions made in the structural studies for tractability, such as no private insurance or no borrowing (Hansen and İmrohorođlu 1992, Hopenhayn and Nicolini 1997).

Equation (27) is not directly implementable because the gap in marginal utilities must be recovered from choice data. The recent literature has proposed the use of three types of choice data to recover the marginal utility gap: consumption (Gruber 1997), liquidity and substitution effects in effort (Chetty 2008a), and reservation wages (Shimer and Werning 2007).

Gruber (1997). Taking a quadratic approximation to the utility function, Gruber observes that

$$\frac{u'(c_l) - u'(c_h)}{u'(c_l)} = \gamma \frac{\Delta c}{c_h}(b) \quad (28)$$

where $\gamma = -\frac{u''(c_h)}{u'(c_h)}c_h$ is the coefficient of relative risk aversion evaluated at c_h and $\Delta c = c_h - c_l$. Gruber posits that the effect of UI benefits on consumption is linear (an assumption that should ideally be evaluated using a structural simulation):

$$\frac{\Delta c}{c_h}(b) = \alpha + \beta b$$

In this specification, α measures the consumption-drop that would occur absent government intervention while β measures the slope of the consumption function with respect to the benefit level. Putting this equation together with (28) and (27) yields the following expression for the marginal welfare gain from increasing the benefit level:

$$M_W(b) = (\alpha + \beta b)\gamma - \frac{\varepsilon_{1-e,b}}{e}. \quad (29)$$

Building on work by Hamermesh (1982), Gruber estimates the consumption-smoothing effect of unemployment insurance (UI) benefits by exploiting changes in UI benefit laws across states in the U.S. coupled with panel data on consumption. He estimates $\alpha = 0.24$ and $\beta = -0.28$, and then calibrates the welfare gain from raising UI benefits using estimates of $\varepsilon_{1-e,b}$ from Meyer (1990). He finds that at conventional levels of risk aversion ($\gamma < 2$), increasing the UI benefit level above the levels observed in his data (roughly 50% of the wage) would lead to substantial welfare losses.

Gruber proceeds to solve for the b^* such that $\frac{dW}{db}(b^*) = 0$ in (29), and finds that b^* is close to zero. These calculations of the optimal benefit level assume that $\frac{\Delta c}{c}$ is linear in b and $\gamma(b)$ and $\varepsilon_{1-e,b}(b)$ do not vary with b . This application of the sufficient statistic formula – which is not guided by a structural model – could be very inaccurate, because it uses ad hoc assumptions to make predictions about counterfactuals that are well out-of-sample. Equation (29) should not be used to make statements about global optima unless one can estimate the sufficient statistics for a range of different benefit levels. Lacking such

estimates, a more coherent method of inferring b^* would be to calibrate a structural model to match the sufficient statistics and simulate the optimal b^* in that model, as in Saez (2001).

A difficulty with (29) is that risk aversion (γ) is known to vary substantially across contexts (Rabin 2000, Chetty and Szeidl 2007). Since Gruber’s results are highly sensitive to the assumed value of γ , more recent studies have sought alternative techniques for recovering the gap in marginal utilities that do not require an estimate of γ .

Chetty (2008a). Chetty (2008a) shows that the gap in marginal utilities in (27) can be backed out from the comparative statics of effort choice. To see this, observe that the first order condition for effort is

$$\psi'(e) = u(c_h) - u(c_l). \quad (30)$$

Now consider the effect of an exogenous cash grant (such as a severance payment to job losers) on effort, holding fixed the private insurance level b_p and the UI tax t :

$$\partial e / \partial A = \{u'(c_h) - u'(c_l)\} / \psi''(e) \leq 0 \quad (31)$$

The effect of increasing the benefit level on effort (again holding b_p and t fixed) is:

$$\partial e / \partial b = -u'(c_l) / \psi''(e) \quad (32)$$

Combining (31) and (32), we see that the ratio of the “liquidity” effect ($\partial e / \partial A$) to the “substitution” effect ($\partial e / \partial w_h = \partial e / \partial A - \partial e / \partial b$) recovers the gap in marginal utilities:

$$\frac{u'(c_l) - u'(c_h)}{u'(c_h)} = \frac{-\partial e / \partial A}{\partial e / \partial A - \partial e / \partial b}$$

Plugging this into (27) yields the following expression for the welfare gain from increasing the benefit level:

$$M_W(b) = \frac{-\partial e / \partial A}{\partial e / \partial A - \partial e / \partial b} - \frac{\varepsilon_{1-e,b}}{e} \quad (33)$$

The intuition for this formula is that the gap between marginal utilities in the good and bad states can be inferred from the extent to which effort is affected by liquidity vs. moral hazard. In a model with perfect consumption smoothing ($c_h = c_l$), the liquidity effect

$\partial e/\partial A = 0$, because a cash grant raises $u(c_h)$ and $u(c_l)$ by the same amount. Chetty shows that like Baily's formula, (33) holds in a general dynamic search model because of envelope conditions in the agent's value functions.

An issue which arises in empirical implementation of (33) is that $\frac{\partial e}{\partial b}$ must be measured holding the tax t fixed, whereas the elasticity $\varepsilon_{1-e,b}$ must be measured while permitting t to vary. Instead of attempting to estimate both parameters, Chetty uses numerical simulations to show that the effect of a UI benefit increase on job finding rates is virtually identical whether or not UI taxes are held fixed. This is because the fraction of unemployed individuals is quite small, making UI tax rates very low.

Chetty implements (33) by estimating the effects of unemployment benefits and severance payments on search intensity using hazard models for unemployment durations. He finds that the welfare gains from raising the unemployment benefit level are small but positive, suggesting that the current benefit level is slightly below the optimum given concavity of $W(b)$. He concludes that the optimal benefit level is close to the current wage replacement rate of approximately 50%.

Shimer and Werning (2007). Shimer and Werning (2007) infer the gap in marginal utilities from the comparative statics of reservation wages instead of effort in a model of job search. They consider a model where the probability of finding a job, e , is determined by the agent's decision to accept or reject a wage offer rather than by search intensity. Wage offers are drawn from a distribution $F(w)$. If the agent rejects the job offer, he receives income of $w_l + b$ as in the model above. For simplicity, assume that the agent has no private insurance ($q = \infty$); allowing $q < \infty$ complicates the algebra but does not affect the final formula. The remainder of the model is specified as in (26).

The agent rejects any net-of-tax wage offer $w - t$ below his outside option $w_l + b$. Therefore, $e = 1 - F(w_l + b + t)$ and the agent's expected value upon job loss is

$$W(b) = eE[u(w - t)|w - t > w_l + b] + (1 - e)u(w_l + b)$$

Note that even though the microeconomic choices of accepting or rejecting wage offers are discrete, the welfare function is smooth because of aggregation, as in (10).

Shimer and Werning’s insight is that $\frac{dW}{db}$ can be calculated using information on the agent’s reservation wage. Suppose we ask the agent what wage he would be willing to accept with certainty *prior* to the start of job search.¹⁸ Define the agent’s reservation wage prior to job search as the wage \bar{w}_0 that would make the agent indifferent about accepting a job immediately to avoid having to take a random draw from the wage offer distribution. The reservation wage \bar{w}_0 satisfies

$$u(\bar{w}_0 - t) = W(b)$$

The government’s problem is to

$$\begin{aligned} \max W(b) &= \max u(\bar{w}_0 - t) \\ &\Rightarrow \max \bar{w}_0 - t \end{aligned} \tag{34}$$

Differentiating (34) gives a sufficient-statistic formula.¹⁹

$$M_W(b) = \frac{d\bar{w}_0}{db} - \frac{dt}{db} = \frac{d\bar{w}_0}{db} - \frac{1-e}{e} \left(1 + \frac{1}{e} \varepsilon_{1-e,b}\right)$$

Intuitively, $\frac{d\bar{w}_0}{db}$ encodes the marginal value of insurance because the agent’s reservation wage directly measures his expected value when unemployed. Shimer and Werning implement (34) using an estimate of $\frac{d\bar{w}_0}{db}$ from Feldstein and Poterba (1984) and find a large, positive value for $M_W(b)$ at current benefit levels. However, they caution that the credibility of existing reservation wage elasticity estimates is questionable, particularly in view of evidence that UI benefit levels have little impact on subsequent wage rates (e.g. Card, Chetty, Weber 2007, van Ours and Vodopivec 2008).

The multiplicity of formulas for $M_W(b)$ illustrates a general feature of the sufficient-

¹⁸Shimer and Werning study a stationary dynamic model with CARA utility where the reservation wage is fixed over time, in which case it does not matter at what point of the spell the reservation wage is elicited.

¹⁹This corresponds to equation (12) in Shimer and Werning (2007), where the unemployment rate is $u = 1 - e$. The slight difference between the formulas (the $\frac{1}{1-u}$ factor in the denominator) arises because Shimer and Werning write the formula in terms of a partial-derivative-based elasticity. Here, $\varepsilon_{1-e,b}$ is the elasticity including the UI tax response needed to balance the budget; in Shimer and Werning’s notation, it is holding the tax fixed.

statistic approach: since the model is not fully identified by the inputs to the formula, there are generally several representations of the formula for welfare gains.²⁰ This flexibility allows the researcher to apply the representation most suitable for his application given the available variation and data. For example, in analyzing disability insurance, it may be easiest to implement Chetty’s (2008a) formula since the available variation permits identification of liquidity and moral hazard effects (Autor and Duggan 2007).

Inefficiencies in Private Insurance. An important assumption made in all three formulas above is that the choices within the private sector are constrained Pareto efficient – that is, total surplus is maximized in the private sector subject to the constraints. In practice, private insurance contracts are likely to be second-best inefficient as well because of adverse selection and moral hazard in private markets. In this case, the envelope condition invoked in deriving (27) is violated because of externalities on the private insurer’s budget constraint that are not taken into account by the individual.

Recent work by Einav, Finkelstein, and Cullen (2008) and Chetty and Saez (2008) identifies sufficient statistics for the welfare gains from social insurance in environments with adverse selection and moral hazard in private insurance markets. Einav et al. develop a method of characterizing the welfare gain from government intervention that uses information about insurance purchase decisions. They show that the demand curve for private insurance and the average cost of providing insurance as a function of the price are together sufficient statistics for welfare analysis. Einav et al. implement their method using quasi-experimental price variation in health insurance policies and find that the welfare gains from government intervention in health insurance markets is small.

Chetty and Saez focus on ex-post behaviors, namely how marginal utilities vary across the high and low states, as in the Baily formula. They develop a simple extension to Gruber’s (1997) implementation of the formula that includes two more parameters – the size of the private insurance market and the crowdout of private insurance by public insurance. Intuitively, the government exacerbates the moral hazard distortion created by private insurance,

²⁰All three formulas hold in models that allow both reservation wage and search intensity choices. Chetty’s (2006a) generalization of Baily’s formula nests the model with stochastic wages. If agents control the arrival rate of offers via search effort, the first order condition for search effort remains the same as in (30), with $Eu'(c_h)$ replacing $u'(c_h)$. It follows that Chetty’s (2008a) formula also holds with stochastic wages.

and must therefore take into account the amount of private insurance and degree of crowdout to calculate the welfare gains from intervention. Chetty and Saez apply their formula to analyze health insurance, and show that naively applying (27) dramatically overstates the welfare gains from government intervention in this case.

These examples illustrate that the sufficient statistic approach can be extended to environments where private sector choices are not second-best efficient. One has to deviate slightly from the general rubric to handle such second-best inefficiencies because the incentive compatibility constraints for private insurance violate Assumption 1. These constraints lead to additional terms in the sufficient statistic formula, increasing the number of moments that need to be estimated for welfare analysis.

V Application 3: Behavioral Models

There is now considerable reduced-form evidence that individuals' behavior deviates systematically from the predictions of neoclassical perfect optimization models; see Table 1 for a few examples and DellaVigna's (2008) review for many more. In light of this evidence, an important new challenge is normative analysis in models where agent's choices deviate from perfect optimization. The budding literature on this topic has proposed some structural approaches to this issue, primarily in the context of time discounting. An early example is Feldstein's (1985) model of social security with myopic agents, in which individuals have a higher discount rate than the social planner. Feldstein numerically calculates the welfare gains from social security policies under various assumptions about the primitives. More recently, a series of papers have used calibrations of Laibson's (1997) β - δ model of hyperbolic discounting to make numerical predictions about optimal policy for agents who are impatient (see Table 1). Another set of studies has modelled the behavioral patterns identified in earlier work – such as “ironing” and “spotlighting” effects in responses to non-linear price schedules – and simulated optimal tax policy in such models (Liebman and Zeckhauser 2004, Feldman and Katuščák 2006).

The difficulty with the structural approach in behavioral applications is that there are often multiple positive models which can explain deviations from rationality, and each of

these models can lead to different welfare predictions. The sufficient statistic approach can be very useful in such situations, because welfare analysis does not require full specification of the positive model underlying observed choices (Bernheim and Rangel 2008). In applications where agents optimize, the main benefit of the sufficient statistic approach is that it simplifies identification. If one had unlimited power to identify primitives, there would be no advantage to using the sufficient statistic approach in such models. In non-optimizing models, however, the sufficient statistic strategy has value even if identification is not a problem because there is no consensus alternative to the neoclassical model.

Given the infancy of this area, there is currently very little work applying sufficient statistic approaches to behavioral models. However, this is a fertile area for further research, as illustrated by Chetty, Looney, and Kroft's (2008) recent analysis of the welfare consequences of taxation when agents optimize imperfectly with respect to taxes. Chetty et al. present evidence that the effect of commodity taxes on demand depend on whether the tax is included in posted prices or not. Taxes that are not included in posted prices – and are hence less salient to consumers – induce smaller demand reductions. There are various psychological and economic theories which could explain why salience affects behavioral responses to taxation, including bounded rationality, forgetfulness, and cue theories of attention. Chetty et al. therefore develop a sufficient-statistic approach to welfare analysis that is robust to specifications of the positive theory of tax salience.

Chetty et al. characterize the efficiency costs and incidence of taxation in a two-good model analogous to the Harberger model presented in section I of this paper. Let x denote the taxed good and y the numeraire. Let demand for x as a function of its pretax price and tax rate be denoted by $x(p, t)$. I assume here that utility is quasilinear in y and production is constant-returns-to-scale. These assumptions simplify the exposition by eliminating income effects and changes in producer prices; Chetty et al. shows that similar results are obtained when these assumptions are dropped.

The agent's true ranking of the consumption bundles (x, y) is described by a smooth,

quasiconcave utility function

$$\begin{aligned} U(x, y) &= u(x) + y \\ &= u(x) + Z - (p + t)x \end{aligned}$$

where the second line imposes that the allocation (x, y) the agent chooses must satisfy the true budget constraint $(p + t)x + y = Z$. Chetty et al. depart from the traditional Harberger analysis by dropping the assumption that the consumption bundle (x, y) is chosen to maximize $U(x, y)$. Instead, they take the demand function $x(p, t)$ as an empirically estimated object generated by a model unknown to the policy maker, permitting $\frac{\partial x}{\partial p} \neq \frac{\partial x}{\partial t}$.

To calculate excess burden, assume that tax revenue is returned to the agent as a lump-sum. Then, under the assumption that individual's utility is a function purely of their ultimate consumption bundle, social welfare is given by

$$W(p, t) = \{u(x) + Z - (p + t)x\} + T(t)$$

where $T(t) = tx(t)$. In non-optimizing models, one must deviate from step 2 of the rubric in section II at this point because the envelope condition used to derive (14) does not hold. Instead, totally differentiate the social welfare function to obtain

$$\frac{dW}{dt} = [u'(x) - p] \frac{dx}{dt} \tag{35}$$

This marginal-utility based expression captures a simple intuition. An infinitesimal tax increase reduces consumption of x by $\frac{dx}{dt}$. The loss in surplus from this reduction in consumption of x is given by the different between willingness to pay for x ($u'(x)$) and the cost of producing x , which equals the price p under the constant-returns-to-scale assumption.

Equation (35) applies in any model, irrespective of how $x(p, t)$ is chosen. Given that $\frac{dx}{dt}$ can be estimated empirically, the challenge in calculating $\frac{dW}{dt}$ – which reflects the main challenge in behavioral public economics more generally – is the recovery of the true preferences ($u'(x)$). In neoclassical models, we use the optimality condition $u'(x) = p + t$ to recover marginal utility and immediately obtain the Harberger formula in (4). Since we

do not know how x is chosen, we cannot use this condition here. Chetty et al. tackle this problem by making the following assumption.

Assumption 2 When tax inclusive prices are fully salient, the agent chooses the same allocation as a fully-optimizing agent:

$$x(p, 0) = \arg \max_x u(x) + Z - px$$

This assumption requires that the agent only makes mistakes with respect to taxes, and not fully salient prices. To see why this assumption suffices to calculate welfare, let $P(x) = x^{-1}(p, 0)$ denote the agent's inverse-price-demand curve. Assumption 2 implies that $P(x) = u'(x)$ via the first order condition for $x(p, 0)$. Plugging this into (35) yields

$$\frac{dW}{dt} = [P(x) - p] \frac{dx}{dt}$$

This formula for $\frac{dW}{dt}$ can be implemented using an estimate of the inverse-price-demand curve $P(x)$. To simplify implementation, Chetty et al. make the approximation that demand $x(p, t)$ is linear in both arguments to obtain

$$\begin{aligned} \frac{dW}{dt} &= \left[\frac{dp}{dx} \cdot (x(p, t) - x(p, 0)) \right] \frac{dx}{dt} \\ &= \left[\frac{dp}{dx} \cdot \frac{dx}{dt} t \right] \frac{dx}{dt} = t\theta \frac{dx}{dt} \end{aligned} \tag{36}$$

where $\theta = \frac{dx}{dt} / \frac{dx}{dp}$ measures the degree of underreaction to the tax relative to the price. This expression, which nests the Harberger formula as the case where $\theta = 1$, shows that the price and tax elasticities of demand are together sufficient statistics to calculate excess burden in behavioral economics models. Intuitively, the tax-demand curve ($\frac{dx}{dt}$) is used to determine the actual effect of the tax on behavior. Then, the price-demand curve ($\frac{dx}{dp}$) is used to calculate the effect of that change in behavior on welfare. The price-demand curve can be used to recover the agent's preferences and calculate welfare changes because it is (by assumption) generated by optimizing behavior.

Because it does not rely on a specific structural model, (36) accommodates *all* errors in

optimization with respect to taxes, and is hence easily adaptable to other applications. For example, confusion between average and marginal income tax rates (Liebman and Zeckhauser (2004), Feldman and Katusčák (2006)) or overestimation of estate tax rates (Slemrod 2006) can be handled using the same formula, without requiring knowledge of individuals' tax perceptions or rules of thumb. Any such behaviors are reflected in the empirically observed tax and price elasticities.

In sum, one can make progress in behavioral welfare economics by making assumptions that narrow the class of models under consideration without fully specifying one particular model. One is effectively forced to make stronger assumptions about the class of models in exchange for relaxing the full optimization assumption. These stronger assumptions make it especially important to implement step 6 of the rubric (model evaluation) in behavioral models. For instance, identifying the structural reasons for why tax salience matters would cast light on the plausibility of Assumption 2.

VI Conclusion: Other Applications

The literature reviewed in this paper has focused on identifying sufficient statistics for normative (welfare) analysis. Sufficient statistics can also be used to answer positive (descriptive) questions. A simple example is predicting the effect of a tax change on tax revenue. One only needs to estimate the elasticity of equilibrium quantity with respect to the tax rate to answer this question. Another example is the literature on capitalization effects (e.g., Summers 1981, Roback 1982, Greenstone and Gallagher 2008), which shows that changes in asset prices are sufficient statistics for distributional incidence in dynamic equilibrium models. The techniques relevant for positive analysis differ from those discussed in this article, as one cannot exploit envelope conditions in most positive applications. However, the general concept is still to formulate answers to questions in terms of a few elasticities instead of a full primitive structure.

Although the sufficient statistic approach has been most widely applied in the public economics literature, it is natural to apply this approach in other areas. I conclude by briefly discussing applications in other fields.

Macroeconomics. A central debate in macroeconomics is whether households adhere to the permanent income hypothesis. The structural approach to answering this question, taken for example by Scholz et al. (2006), is to specify a dynamic model of optimization and test whether observed consumption and savings patterns are consistent with those predicted by the model. The sufficient statistic counterpart to this approach is to isolate one moment – such as the drop in consumption at retirement or the sensitivity of behavior to cash-on-hand – that is adequate to test between models (see e.g., Aguiar and Hurst 2005; Card, Chetty, and Weber 2007). Similarly, models of business cycles and growth can be distinguished simply by identifying the “labor wedge” (Shimer 2008).

Labor. Labor economists have studied the effects of minimum wages on employment and wages extensively using reduced-form methods. Such evidence can be mapped into statements about optimal policy using a sufficient statistic approach (Lee and Saez 2008). Another potential application is to the analysis of returns to schooling. Many studies have investigated the effects of schooling on behaviors such as job mobility and occupation choice. A sufficient statistic approach would intuitively suggest that examining effects on total earnings is adequate to measure the benefits of additional schooling in a model where agents optimize.

Development. Starting with Townsend (1994), a large literature in development has studied risk sharing arrangements. While it is informative to understand the mechanisms through which shocks are smoothed, identifying consumption fluctuations and risk aversion is sufficient to make inferences about the welfare costs of shocks (Chetty and Looney 2006). Hence, quasi-experimental evidence such as that of Gertler and Gruber’s (2002) study of health shocks in Indonesia sheds light on welfare and optimal policy even though it does not fully characterize the structure of insurance and risk sharing networks. More generally, one may be able to give precise answers to policy questions using estimates from randomized experiments coupled with sufficient statistic formulas derived from standard structural models. For example, the effects of interventions on health, education, or income may encode all that is needed for policy analysis.

Industrial Organization. Weyl and Fabinger (2009) show that estimates of the pass-through of cost shocks are sufficient statistics for questions ranging from the effects of mergers

on markups to the effects of price caps on market structure. Researchers who seek to develop sufficient statistic approaches to other problems in IO must confront two challenges. First, many questions of interest concern discrete changes such as antitrust policy or the introduction of a new product. Second, much of the IO literature focuses on models of strategic interaction rather than price theory. In strategic games, small changes in exogenous parameters can lead to jumps in behavior. Since the techniques used in this paper rely on smoothness, one may need to develop different techniques to apply the sufficient statistic approach to IO problems. These are interesting topics for further research at the interface of structural and reduced-form methods.

References

- Aguiar, Mark and Hurst, Erik (2005). "Consumption vs. Expenditure," *Journal of Political Economy*, 113(5), 919–948.
- Amador, Manuel, Iván Werning, and George-Marios Angeletos (2006). "Commitment vs. Flexibility," *Econometrica*, 74(2): 365–396.
- Anderson, Patricia M., and Bruce D. Meyer (1997). "Unemployment Insurance Takeup Rates and the After-Tax Value of Benefits," *Quarterly Journal of Economics* CXII: 913–937.
- Angeletos, George-Marios, David Laibson, Andrea Repetto, Jeremy Tobacman, and Stephen Weinberg (2001). "The Hyperbolic Consumption Model: Calibration, Simulation, and Empirical Evaluation," *Journal of Economic Perspectives*, 47–68.
- Angrist, Joshua D. and Alan B. Krueger (1991). "Does Compulsory School Attendance Affect Schooling and Earnings?" *Quarterly Journal of Economics*, 106(4): 979–1014.
- Angrist, Joshua D., Kathryn Graddy, and Guido Imbens (2000). "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish," *Review of Economic Studies*, 67(3): 499–527.
- Ashraf, Nava, Dean Karlan, and Wesley Yin (2006). "Tying Odysseus to the Mast: Evidence from a Commitment Savings Product in the Philippines," *Quarterly Journal of Economics*, 121(2): 635–672.
- Auerbach, Alan J. (1985). "The Theory of Excess Burden and Optimal Taxation." In Vol. 1, *Handbook of Public Economics*, ed. Alan J. Auerbach and Martin S. Feldstein, 67–127. Amsterdam: Elsevier Science Publishers.
- Autor, David H. and Mark G. Duggan (2003). "The Rise In The Disability Rolls And The Decline In Unemployment," *Quarterly Journal of Economics*, 118(1): 157–205.
- Autor, David H. and Mark G. Duggan (2007). "Distinguishing Income from Substitution Effects in Disability Insurance," *American Economic Review Papers and Proceedings*, 97(2): 119–124.
- Baily, Martin N. (1978). "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, 10 (December): 379–402.
- Ballard, Charles L., John B. Shoven, and John Whalley (1985). "The total welfare cost of the United States tax system: A general equilibrium approach," *National Tax Journal*, 38(2): 125–140.
- Bernheim, B. Douglas and Antonio Rangel (2008). "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics," *Quarterly Journal of Economics*, forthcoming.
- Berry, Steven (1994). "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics* 25(2): 242–262.
- Berry, Steven, James Levinsohn, and Ariel Pakes (1995). "Automobile Prices in Market Equilibrium," *Econometrica* 63(4): 841–90.
- Blau, Francine D. and Lawrence M. Kahn (2007). "Changes in the Labor Supply Behavior of Married Women: 1980–2000," *Journal of Labor Economics* 25: 393–438.
- Blundell, Richard, Duncan, Alan and Meghir, Costas (1998). "Estimating Labor Supply Responses Using Tax Reforms." *Econometrica* 66(7): 827–862.
- Blundell, Richard, Alan Duncan, Julian McCrae, and Costas Meghir (2000). "The labour market impact of the working families' tax credit," *Fiscal Studies*, 21(1): 75–103.

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008). “Consumption inequality and partial insurance,” *American Economic Review*, forthcoming.

Card, David (1990). “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review*, 43, 245–257.

Card, David, Raj Chetty, and Andrea Weber (2007). “Cash-on-Hand and Competing Models of Intertemporal Behavior: New Evidence from the Labor Market,” *Quarterly Journal of Economics*, 122(4): 1511–1560.

Chetty, Raj (2006a). “A General Formula for the Optimal Level of Social Insurance,” *Journal of Public Economics*, 90 (November): 1879–1901.

Chetty, Raj (2006b). “A New Method of Estimating Risk Aversion,” *American Economic Review*. 96(5): 1821–1834.

Chetty, Raj and Adam Looney (2006). “Consumption Smoothing and the Welfare Consequences of Social Insurance in Developing Economies,” *Journal of Public Economics*, 90: 2351–2356.

Chetty, Raj and Adam Szeidl (2007). “Consumption Commitments and Risk Preferences,” *Quarterly Journal of Economics*, 122(2): 831–877.

Chetty, Raj (2008a). “Moral Hazard vs. Liquidity and Optimal Unemployment Insurance,” *Journal of Political Economy*, 116(2): 173–234.

Chetty, Raj (2008b). “Is the Taxable Income Elasticity Sufficient to Calculate Deadweight Loss? The Implications of Evasion and Avoidance,” *American Economic Journal - Economic Policy*, forthcoming.

Chetty, Raj, Adam Looney, and Kory Kroft (2008). “Salience and Taxation: Theory and Evidence,” *American Economic Review*, forthcoming.

Chetty, Raj and Emmanuel Saez (2008). “Optimal Taxation and Social Insurance with Endogenous Private Insurance.” NBER Working Paper No. 14403.

Chetty, Raj and Emmanuel Saez (2009). “Teaching the Tax Code: Earnings Responses to an Experiment with EITC Recipients,” NBER Working Paper No. 14836 .

Dawkins, Christina, T.N. Srinivasan, and John Whalley (2001). “Calibration,” in *Handbook of Econometrics*, Vol. 5, J. Heckman and E. Leamer (eds.), North Holland, Amsterdam: 3653–3703.

Deaton, Angus and John Muellbauer (1980). “An Almost Ideal Demand System,” *American Economic Review*, 70(3): 312–26.

Deaton, Angus (2009). “Instruments of Development: Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development,” NBER Working Paper No. 14690.

DellaVigna, Stefano (2008). “Psychology and Economics: Evidence from the Field,” *Journal of Economic Literature*, forthcoming.

DellaVigna, Stefano and M. Daniele Paserman (2005). “Job Search and Impatience,” *Journal of Labor Economics*, 23: 527–588.

Diamond, Peter A. (1998). “Optimal Income Taxation: An Example with a U-Shaped Pattern of Optimal Marginal Tax Rates,” *American Economic Review*, 88(1): 83–95.

Einav, Liran, Amy Finkelstein, and Mark Cullen (2008). “Estimating Welfare in Insurance Markets Using Variation in Prices,” NBER Working Paper No. 14414.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf (2008). “The Welfare Cost of Asymmetric Information: Evidence from the U.K. Annuity Market,” NBER Working Paper No.

W13228.

Eissa, Nada and Jeffrey B. Liebman (1996). "Labor Supply Response to the Earned Income Tax Credit," *Quarterly Journal of Economics*, 111(2): 605–637.

Feldman, Naomi E. and Peter Katuščák (2006). "Should the Average Tax Rate be Marginalized?" CERGE Working Paper No. 304.

Feldstein, Martin S. and James Poterba (1984). "Unemployment insurance and reservation wage," *Journal of Public Economics*, 23(1-2): 141–167.

Feldstein, Martin S. (1985) "The Optimal Level of Social Security Benefits," *Quarterly Journal of Economics*, 100(2): 303–20.

Feldstein, Martin S. (1995). "The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act," *Journal of Political Economy*, 103(3): 551–572.

Feldstein, Martin S. (1999). "Tax Avoidance and the Deadweight Loss of the Income Tax," *Review of Economics and Statistics*, 81(4): 674–680.

Finkelstein, Amy (2007). "The Aggregate Effects of Health Insurance: Evidence from the Introduction of Medicare," *Quarterly Journal of Economics* 122(1): 1–37.

Genesove, David and Christopher Mayer (2001). "Loss Aversion And Seller Behavior: Evidence From The Housing Market," *Quarterly Journal of Economics*, 116(4): 1233-1260.

Gertler, Paul J. and Jonathan Gruber (2002). "Insuring Consumption Against Illness," *American Economic Review*, 92(1): 51–70.

Golosov, Mikhail and Aleh Tsyvinski (2006). "Designing Optimal Disability Insurance: A Case for Asset Testing," *Journal of Political Economy*, 114(2): 257–269.

Golosov, Mikhail and Aleh Tsyvinski (2007). "Optimal Taxation with Endogenous Insurance Markets," *Quarterly Journal of Economics*, 122(2): 487–534.

Goolsbee, Austan (2000). "What Happens When You Tax the Rich? Evidence from Executive Compensation," *Journal of Political Economy*, 108(2): 352–378.

Gorodnichenko, Yuriy, Jorge Martinez-Vazquez, and Klara Sabirianova Peter (2009). "Myth and Reality of Flat Tax Reform: Micro Estimates of Tax Evasion and Productivity Response in Russia," *Journal of Political Economy*, forthcoming.

Goulder, Lawrence H. and Roberton C. Williams (2003). "The Substantial Bias from Ignoring General Equilibrium Effects in Estimating Excess Burden, and a Practical Solution," *Journal of Political Economy*, 111(4): 898–927.

Greenstone, Michael and Justin Gallagher (2008). "Does Hazardous Waste Matter? Evidence from the Housing Market and the Superfund Program," *Quarterly Journal of Economics*, 123(3): 951–1003.

Gruber, Jonathan (1997). "The Consumption Smoothing Benefits of Unemployment Insurance," *American Economic Review*, 87 (March): 192–205.

Gruber, Jonathan and David Wise (1999). *Social security and retirement around the world*. Chicago: University of Chicago Press.

Gruber, Jonathan and Emmanuel Saez (2002). "The Elasticity of Taxable Income: Evidence and Implications," *Journal of Public Economics*, 84: 1–32.

Hamermesh, Daniel S. (1982). "Social Insurance and Consumption: An Empirical Inquiry," *American Economic Review*, 72(1): 101–113.

Hansen, Gary and Ayse İmrohoroğlu (1992). "The Role of Unemployment Insurance in an Economy with Liquidity Constraints and Moral Hazard," *Journal of Political Economy*, 100 (January): 118–142.

- Harberger, Arnold C. (1964). "The Measurement of Waste," *American Economic Review*, 54(3): 58–76.
- Hausman Jerry A. (1981). "Exact Consumer's Surplus and Deadweight Loss," *American Economic Review*, 71(4): 662–676.
- Hausman, Jerry A. and Whitney K. Newey (1995). "Nonparametric Estimation of Exact Consumers Surplus and Deadweight Loss," *Econometrica*, 63(6): 1445–1476.
- Heckman, James H. and Edward Vytlacil (2001). "Policy Relevant Treatment Effects." *American Economic Review*, 91(2): 107–111.
- Heckman, James H. and Edward Vytlacil (2005). "Structural Equations, Treatment Effects, and Econometric Policy Evaluation." *Econometrica*, 73(3): 669–738.
- Heckman, James J. and Edward Vytlacil (2007). "Econometric Evaluation of Social Programs," in *Handbook of Econometrics*, Vol 6, J. Heckman and E. Leamer (eds.), North Holland: Amsterdam: 4779–4874.
- Hines, James R., Jr. (1999). "Three sides of Harberger triangles," *Journal of Economic Perspectives*, 13(2): 167–188.
- Hopenhayn, Hugo and Juan Pablo Nicolini (1997). "Optimal Unemployment Insurance," *Journal of Political Economy*, 105 (April): 412–438.
- Hoynes, Hilary W. (1996). "Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation Under the AFDC-UP Program," *Econometrica*, 64(2): 295–332.
- Imbens, Guido (2009). "Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009)," NBER Working Paper No. 14896.
- Imbens, Guido and Jeffrey Wooldridge (2008). "Recent Developments in the Econometrics of Program Evaluation," NBER Working Paper No. 14251.
- İmrohoroğlu, Ayşe, Selahattin İmrohoroğlu, and Douglas Joines (2003). "Time Inconsistent Preferences and Social Security," *Quarterly Journal of Economics* 118(2): 745–784.
- Joint Economic Committee, U.S. Congress (2001). "Economic Benefits of Personal Income Tax Rate Reductions."
- Keane, M. and K. Wolpin (1997). "The Career Decisions of Young Men." *Journal of Political Economy*, 105(3): 473–522.
- Keane, M., and R. Moffitt (1998). "A Structural Model of Multiple Welfare Program Participation and Labour Supply," *International Economic Review*, 39: 553–89
- Koopmans, Tjalling C. (1953). "Identification Problems in Economic Model Construction," in *Studies of Econometric Method*, ed. William C. Hood and Tjalling C. Koopmans. Cowles Commission for Research in Economics, Monograph No. 14, pp 27–48. John Wiley & Sons, Inc., New York.
- Laibson, David (1997). "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 62, May, 443–77.
- Lalive, Rafel, Jan van Ours, and Josef Zweimuller (2006). "How Changes in Financial Incentives Affect the Duration of Unemployment," *Review of Economic Studies* 73(4): 1009–1038.
- Lalonde, Robert J. (1986). "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," *American Economic Review*, 76, 604–620.
- Lee, David and Emmanuel Saez (2008). "Optimal Minimum Wage in Competitive Labor Markets," NBER Working Paper No. 14320.

- Lentz, Rasmus (2009). “Optimal Unemployment Insurance in an Estimated Job Search Model with Savings,” *Review of Economic Dynamics*, 12(1):37–57.
- Levitt, Steven D. (1997). “Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime,” *American Economic Review*, 87(3): 270–90.
- Liebman, Jeffrey B. and Richard J. Zeckhauser (2004). “Schmeduling,” Harvard KSG Working Paper.
- Lumsdaine, Robin, James Stock, and David Wise (1992). “Three models of retirement: computational complexity versus predictive ability.” In Wise, D., Ed. *Topics in the Economics of Aging*, pp. 19–60. University of Chicago Press, Chicago.
- Madrian, Brigitte C. and Dennis F. Shea (2001). “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior,” *Quarterly Journal of Economics*, 116(4): 1149–1187.
- Marschak, Jacob (1953). “Economic Measurements for Policy and Prediction,” in *Studies of Econometric Method*, ed. William C. Hood and Tjalling C. Koopmans. Cowles Commission for Research in Economics, Monograph No. 14, pp. 1–26. John Wiley & Sons, Inc., New York.
- Meyer, Bruce D. (1990). “Unemployment Insurance and Unemployment Spells,” *Econometrica*, 58: 757–782.
- Meyer, Bruce D. and Dan T. Rosenbaum (2001). “Welfare, The Earned Income Tax Credit, And The Labor Supply Of Single Mothers,” *Quarterly Journal of Economics*, 116(3): 1063–1114.
- Mirrlees, James A. (1971). “An Exploration in the Theory of Optimum Income Taxation,” *The Review of Economic Studies*, 38(2): 175–208.
- Piketty, Thomas (1997). “La Redistribution fiscale contre le chômage,” *Revue française d’économie*, 12(1): 157–203.
- Rabin, Matthew (2000). “Risk Aversion and Expected-Utility Theory: A Calibration Theorem,” *Econometrica*, 68(5): 1281–1292.
- Roback, Jennifer (1982). “Wages, Rents, and the Quality of Life.” *Journal of Political Economy* 90: 1257–1278.
- Rosenzweig, Mark R. and Kenneth I. Wolpin (2000). “Natural ‘Natural Experiments’ in Economics.” *Journal of Economic Literature* 38(4): 827–874.
- Rust, John and Christopher Phelan (1997). “How Social Security and Medicare Affect Retirement Behavior in a World of Incomplete Markets.” *Econometrica* 65: 781–832.
- Saez, Emmanuel (2001). “Using Elasticities to Derive Optimal Income Tax Rates,” *Review of Economic Studies*, 68: 205–229.
- Scholz, Karl, Ananth Seshadri, and Surachai Khittrakun (2006). “Are Americans Saving ‘Optimally’ for Retirement?” *Journal of Political Economy*, 114(4): 607–643.
- Shapiro, Jesse M. (2005). “Is there a daily discount rate? Evidence from the food stamp nutrition cycle,” *Journal of Public Economics* 89(2-3): 303–325.
- Shimer, Robert and Iván Werning (2007). “Reservation Wages and Unemployment Insurance,” *Quarterly Journal of Economics*, 122(3): 1145–1185.
- Shimer, Robert (2008). “Convergence in Macroeconomics: The Labor Wedge” Forthcoming, *American Economic Journal - Macroeconomics*, 1(1): 280–97.
- Shoven, John B. (1976). “The incidence and efficiency effects of taxes on income from capital,” *Journal of Political Economy*, 84(6): 1261–1283.

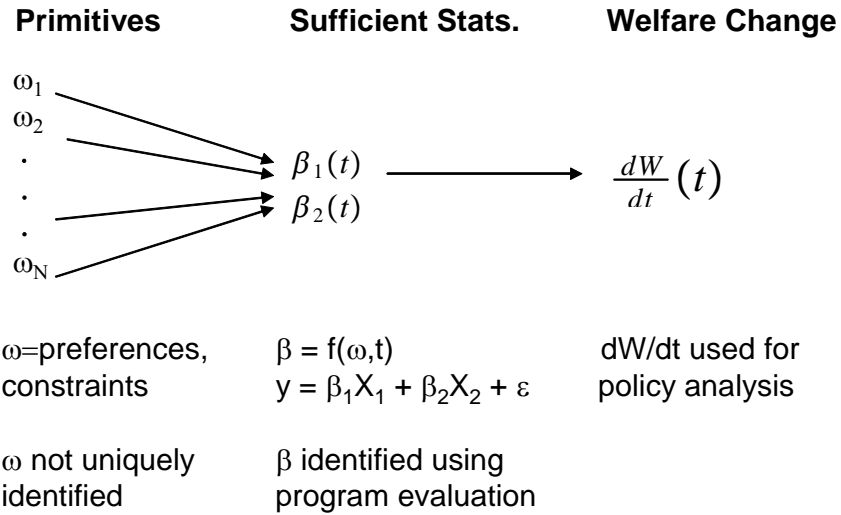
- Slemrod, Joel B. (1995). “Income Creation or Income Shifting? Behavioral Responses to the Tax Reform Act of 1986,” *American Economic Review*, 85(2): 175–180.
- Slemrod, Joel B. (2006). “The Role of Misconceptions in Support for Regressive Tax Reform,” *National Tax Journal*, 59(1): 57–75.
- Small, Kenneth A. and Harvey S. Rosen (1981). “Applied Welfare Economics with Discrete Choice Models,” *Econometrica* 49(1): 105–30.
- Stone, Richard (1954). “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand,” *The Economic Journal*, 64(255): 511–527.
- Summers, Larry (1981). “Taxation and Corporate Investment: A q-Theory Approach.” *Brookings Papers on Economic Activity* 1981(1): 67–127.
- Townsend, Robert (1994). “Risk and insurance in village India.” *Econometrica* 62 (3): 539–591.
- Train, Kenneth (2003). *Discrete Choice Methods with Simulation*, Cambridge: Cambridge University Press.
- Tuomala, Matti (1990). *Optimal Income Tax and Redistribution*, Oxford: Clarendon Press.
- van Ours, Jan C. and Milan Vodopivec (2008). “Shortening the Potential Duration of Unemployment Benefits Does Not Affect the Quality of Post-Unemployment Jobs: Evidence from a Natural Experiment,” *Journal of Public Economics*, 92 (3-4): 684–695
- Weinzierl, Matthew (2008). “The Surprising Power of Age-Dependent Taxes,” Harvard University working paper.
- Weyl, Glen E. and Michal Fabinger (2009). “Pass-through as an Economic Tool,” Harvard University working paper.
- Wolpin, Kenneth I. (1987). “Estimating a Structural Search Model: The Transition from School to Work,” *Econometrica*, 55(4): 801–817.

TABLE 1
Recent Examples of Structural, Reduced-Form, and Sufficient Statistic Studies

	Structural	Reduced Form	Sufficient Statistic
Taxation	Hoynes (1996) Keane and Moffitt (1998) Blundell et al. (2000) Golosov and Tsyvinski (2007) Weinzierl (2008)	Eissa and Liebman (1996) Blundell, Duncan, and Meghir (1998) Goolsbee (2000) Meyer and Rosenbaum (2001) Blau and Khan (2007)	Feldstein (1999) Diamond (1998) Saez (2001) Goulder and Williams (2003) Chetty (2008b)
Social Insurance	Rust and Phelan (1997) Golosov and Tsyvinski (2006) Lentz (2009) Blundell, Pistaferri, and Preston (2008) Einav, Finkelstein, and Schrimpf (2008)	Anderson and Meyer (1997) Gruber and Wise (1999) Autor and Duggan (2003) Lalive, van Ours, and Zweimuller (2006) Finkelstein (2007)	Gruber (1997) Chetty (2006a) Shimer and Werning (2007) Chetty (2008a) Einav, Finkelstein, and Cullen (2008)
Behavioral Models	Angeletos et al. (2001) İmrohoroğlu, İmrohoroğlu, and Joines (2003) Liebman and Zeckhauser (2004) DellaVigna and Paserman (2005) Amador, Werning, and Angeletos (2006)	Genesove and Mayer (2001) Madrian and Shea (2001) Shapiro (2005) Ashraf, Karlan, and Yin (2006) Chetty and Saez (2009)	Bernheim and Rangel (2008) Chetty, Looney, and Kroft (2008)

Notes: Categories used to classify papers are defined as follows. Structural: estimate or calibrate primitives to make predictions about welfare. Reduced form: estimate high-level behavioral elasticities qualitatively relevant for policy analysis, but do not provide quantitative welfare results. Sufficient statistic: make predictions about welfare without estimating or specifying primitives. This list includes only selected examples that relate to the topics discussed in the text, and omits many important contributions in each category.

FIGURE 1
THE SUFFICIENT STATISTIC APPROACH



NOTE—Consider a policy instrument t that affects social welfare $W(t)$. The structural approach maps the primitives (ω) directly to the effects of the policy on welfare ($\frac{dW}{dt}$). The sufficient statistic approach leaves ω unidentified and instead identifies a smaller set of high-level parameters (β) using program evaluation methods, e.g. via a regression of an outcome y on exogenous variables X . The β vector is “sufficient” for welfare analysis in that any vector ω consistent with β implies the same value of $\frac{dW}{dt}$. Identifying β does not identify ω because there are multiple ω vectors consistent with a single β vector.