AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Pat Kline & Chris Walters

January 6-7, 2021

This course covers recent developments in empirical labor economics along with econometric methods commonly used in contemporary research on labor markets.

Course Outline

Part I: Education and Human Capital

Returns to schooling

Card, D. (1999). “The causal effect of education on earnings.” Handbook of Labor Economics, Volume 3A.

Angrist, J., and Krueger, A. (1991). “Does compulsory schooling attendance affect schooling and earnings?” *Quarterly Journal of Economics* 106 (4).

Zimmerman, S. (2014). “The returns to college admission for academically marginal students." *Journal of Labor Economics* 32 (4).

Clark, D., and P. Martorell (2014). “The signaling value of a high school diploma.” *Journal of Political Economy* 122 (2).

Methods: Potential outcomes, instrumental variables (IV), regression discontinuity (RD)

College quality

Dale, S., and A. Krueger (2002). “Estimating the payoff to attending a more selective college: An application of selection on observables and unobservables.” *Quarterly Journal of Economics* 117 (4).

Mountjoy, J., and B. Hickman (2020). “The returns to college(s): estimating value-added and match effects in higher education.” Working paper.

Methods: Matching/selection on observables

Part II: Self-selection

Roy and generalized Roy models

Roy, A. (1951). “Some thoughts on the distribution of earnings.” *Oxford Economic Papers New Series* 3.

Eisenhauer, P., James Heckman, J., and Vytlacil, E. (2015). “The generalized Roy model and the cost-benefit analysis of social programs.” *Journal of Political Economy* 123 (2).

Kline, P., and Walters, C. (2019). “On Heckits, LATE, and numerical equivalence.” *Econometrica* 87 (2).

Methods: Control functions, IV, marginal treatment effects

Self-selection applications

Kline, P., and Walters, C. (2016). “Evaluating public programs with close substitutes: the case of Head Start.” *Quarterly Journal of Economics* 131(4).

Kirkeboen, L., Leuven, E., and Mogstad, M. (2016). “Field of study, earnings, and self-selection.” *Quarterly Journal of Economics* 131(3).

Walters, C. (2018). “The demand for effective charter schools.” *Journal of Political Economy* 126 (6).

Methods: IV, discrete choice, dynamic models

Part III: Discrimination

Group differences in the labor market

Altonji, J., and R. Blank (1999). “Race and gender in the labor market.” Handbook of Labor Economics, Volume 3.

Guryan, J., and Charles, K. (2013). “Taste-based or statistical discrimination: the economics of discrimination returns to its roots.” *The Economic Journal* 123 (572).

Chetty, R., N. Hendren, M. Jones, and S. Porter (2020). “Race and economic opportunity in the United States: an intergenerational perspective.” *Quarterly Journal of Economics* 135 (2).

Methods: Oaxaca-Blinder decompositions

Experimental approaches

Bertrand, M., and S. Mullainathan (2004). “Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination.” *American Economic Review* 94 (4).

Kline, P., and Walters, C. (forthcoming). “Reasonable doubt: experimental detection of job-level employment discrimination.” *Econometrica*.

Methods: Experiments, empirical Bayes, bounds

Algorithmic bias

Rose, E. K. (forthcoming). “Who gets a second chance? Effectiveness and equity in supervision of criminal offenders.” *Quarterly Journal of Economics*.

Kleinberg, J., H. Lakkaraju, J. Leskovec, J. Ludwig, and S. Mullainathan (2018). “Human decisions and machine predictions.” *Quarterly Journal of Economics* 133 (1).

Methods: Difference-in-differences

Part IV: Minimum Wages

Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. *American Economic Review*, *109*(8), 2766-94.

Card, D., and Krueger, A. (1994). “Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania.” *American Economic Review* 84 (4).

Cengiz, D., Dube, A., Lindner, A., and Zipperer, B. (2019). “The effect of minimum wages on low-wage jobs.” *Quarterly Journal of Economics* 134 (3).

Dustmann, C., Lindner, A., Umkehrer, M., & vom Berge, P. (2020). *Reallocation Effects of the Minimum Wage* (No. 2007). Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London.

Harasztosi, P., and Lindner, A. (2019). “Who pays for the minimum wage?” *American Economic Review* 109 (8).

Giuliano, L. (2013). Minimum wage effects on employment, substitution, and the teenage labor supply: Evidence from personnel data. *Journal of Labor Economics*, *31*(1), 155-194.

Methods: Difference-in-differences (DiD), differential exposure designs

Part V: Firm Wage Premia

Abowd, J., Kramarz, F., and Margolis, D. (1999). “High wage workers and high wage firms.” *Econometrica* 67 (2).

Card, D., Heining, J., and Kline, P. (2013). “Workplace heterogeneity and the rise of West German wage inequality.” *Quarterly Journal of Economics* 128 (3).

Card, D., Cardoso, A., and Kline, P. (2016). “Bargaining, sorting, and the gender wage gap: quantifying the impact of firms on the relative pay of women.” *Quarterly Journal of Economics* 131 (2).

Gerard, F., Lagos, L., Severnini, E., & Card, D. (2018). *Assortative matching or exclusionary hiring? the impact of firm policies on racial wage differences in brazil* (No. w25176). National Bureau of Economic Research.

Goldschmidt, D., and Schmieder, J. (2017). “The rise of domestic outsourcing and the evolution of the German wage structure.” *Quarterly Journal of Economics* 132 (3).

Kline, P., Saggio, R., and Sølvsten, M. (2020). “Leave-out estimation of variance components.” *Econometrica* 88 (5).

Lachowska, M., Mas, A., Saggio, R. D., & Woodbury, S. A. (2020). *Do firm effects drift? Evidence from Washington administrative data* (No. w26653). National Bureau of Economic Research.

Methods: High dimensional fixed effects, Oaxaca decomps

Part VI: Monopsony, Rent-Sharing, and Outside Options

Azar, J., Berry, S., and Marinescu, I. (2019). “Estimating labor market power.” Working paper.

Card, D., Cardoso, A., Heining, J., and Kline, P. (2018). “Firms and labor market inequality: evidence and some theory.” *Journal of Labor Economics* 36 (S1).

Di Addario, S., Kline, P., Saggio, R., and Solvsten, M. (2020). “It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages.” IRLE Working Paper #104-20.

Dube, A., Jacobs, J., Naidu, S., and Suri, S. (2020). “Monopsony in online labor markets.” *American Economic Review: Insights* 2(1).

Jäger, S., Schoefer, B., Young, S., and Zweimüller, J. (2020). “Wages and the value of nonemployment.” *Quarterly Journal of Economics* 135 (4).

Kline, P., Petkova, N., Williams, H., and Zidar, O. (2019). “Who profits from patents? Rent-sharing at innovative firms.” *Quarterly Journal of Economics* 134 (3).

Staiger, D., Spetz, J., and Phibbs, C. (2010). “Is there monopsony in the labor market? Evidence from a natural experiment.” *Journal of Labor Economics* 28 (2).

Methods: Discrete choice modeling, DiD, experiments, fixed effects

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture 1 - Education and Human Capital

Chris Walters

University of California, Berkeley and NBER

Labor Economics and Applied Econometrics

- This course covers topics in modern labor economics
- We will also cover econometric tools that are commonly used in contemporary applied microeconomics
- **Instructors:**
  - Pat Kline, pkline@econ.berkeley.edu
  - Chris Walters, crwalters@econ.berkeley.edu
- **Schedule:** January 6-7, 11:15AM-6:45PM (with breaks)
- Syllabus, schedule, and slides available on course website

Human Capital and Education

- First topic: human capital and education
- Motivated by human capital paradigm
  - Worker skills as a form of capital
  - Choose how much to invest in skills, balancing increased earnings in the future against opportunity cost of earnings foregone in the present
  - Key parameter: causal return to schooling
- The causal return to schooling answers a counterfactual question: how much more would a particular person earn if s/he spent more time in school?
- We will discuss such questions in the language of **potential outcomes**

Potential Outcomes

- Consider a person i deciding whether to attend college
- The indicator $D_i \in \{0, 1\}$ takes a value of 1 if i attends college, and 0 otherwise
- $Y_i(1)$ denotes i's potential earnings if she attends college
- $Y_i(0)$ denotes i's potential earnings if she does not attend college
- Potential outcomes are defined by a hypothetical manipulation: what *would happen* to a particular person in one condition or the other
- The causal effect of college on person i's earnings is defined as:

$$\delta_i = Y_i(1) - Y_i(0).$$

- This simple model of causality is called the **Rubin causal model** (Holland 1986)

The Fundamental Problem of Causal Inference

- In the real world, a person either attends college, or she doesn't
- This means only one potential outcome will ever be observed – the other is **counterfactual**
- The observed outcome, Y_i, equals $Y_i(0)$ if $D_i = 0$ and $Y_i(1)$ if $D_i = 1$. We can then write

$$Y_i = Y_i(0) + (Y_i(1) - Y_i(0))D_i$$

- Since we can never observe both $Y_i(0)$ and $Y_i(1)$, we can't see δ_i for any individual. This is known as the **fundamental problem of causal inference**
- The econometric methods we will cover can be viewed as approaches to imputing missing potential outcomes
- We can never hope to recover δ_i for an individual person, but sometimes we can recover certain averages

Average Treatment Effects

- The **average treatment effect** for a population is defined as:

$$ATE = E\left[Y_i(1) - Y_i(0)\right]$$

- "Treatment effects" language is adopted from medical trials
  - $Y_i(1)$ is i's outcome if assigned the treatment (college)
  - $Y_i(0)$ is i's outcome if assigned the control condition (no college)
  - $\delta_i = Y_i(1) - Y_i(0)$ is i's **treatment effect**

- Other treatment effect parameters of interest include the **effect of treatment on the treated** (TOT), and the **effect of treatment on the non-treated** (TNT):

$$TOT = E\left[Y_i(1) - Y_i(0) | D_i = 1\right]$$

$$TNT = E\left[Y_i(1) - Y_i(0) | D_i = 0\right]$$

Treatment Effects and Selection Bias

- Consider a comparison of average observed earnings for individuals that attend college vs. those that don't:

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=E\left[Y_i(1)|D_i=1\right]-E\left[Y_i(0)|D_i=0\right]$$

- Add and subtract $E[Y_i(0)|D_i=1]$ on the right-hand side:

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=E[Y_i(1)-Y_i(0)|D_i=1]$$

$$+E[Y_i(0)|D_i=1]-E\left[Y_i(0)|D_i=0\right]$$

Treatment Effects and Selection Bias

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=\underbrace{E[Y_i(1)-Y_i(0)|D_i=1]}_{TOT}$$

$$+\underbrace{E[Y_i(0)|D_i=1]-E\left[Y_i(0)|D_i=0\right]}_{Selection\ \ Bias}$$

- This expression decomposes the observed treatment/control difference into the TOT plus a **selection bias** term given by the difference in average $Y_i(0)$'s between treatment and control
- Selection bias arises if the observed outcome for the control group fails to match the missing counterfactual for the treatment group

Treatment Effects and Selection Bias

- Note that we could've written

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=\underbrace{E[Y_i(1)-Y_i(0)|D_i=0]}_{TNT}$$

$$+\underbrace{E[Y_i(1)|D_i=1]-E\left[Y_i(1)|D_i=0\right]}_{Selection\ \ Bias}$$

- Here selection bias arises if the observed outcome for the treatment group fails to match the missing counterfactual for the control group
- Definition of selection bias depends on the question we're asking – which counterfactual outcome are we trying to impute?

The RCT Ideal

- Suppose the treatment is assigned independently of potential outcomes:

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i$$

- Then

$$E[Y_i|D_i = 1] - E[Y_i|D_i = 0] = E[Y_i(1)|D_i = 1] - E[Y_i(0)|D_i = 0]$$

$$= E[Y_i(1)] - E[Y_i(0)]$$

$$= ATE.$$

- Assigning treatment randomly as in a **randomized controlled trial** (RCT) guarantees independence of potential outcomes from treatment
  - Randomization eliminates selection bias
  - Implies treatment/control difference $= ATE = TOT = TNT$
- Often the treatment of interest is not randomized. Other **research designs** aim to isolate comparisons that are as good as random

Human Capital Investments

- Return to the idea of human capital investment
- Start with a simple model of schooling investments, as in Card (1999)
- Individual i chooses duration of schooling S to maximize the present discounted value of earnings:

$$\int_{S}^{\infty} e^{-r_i t} Y_i(S) dt$$

- The **potential earnings function** $Y_i(S)$ now describes i's potential earnings for every possible schooling level
- Attending S years of school results in zero earnings until S, and then $Y_i(S)$ thereafter
- Interest rate r_i determines how i discounts future earnings

Optimal Schooling Choice

- Optimal schooling choice maximizes *PDV*:

$$S_i^* = \arg\max_S \int_S^\infty e^{-r_i t} Y_i(S) dt$$

- First-order condition:

$$\frac{Y_i'(S_i^*)}{Y_i(S_i^*)} = r_i$$

- Marginal benefit/marginal cost formula: at any S, can invest current earnings $Y_i(S)$ and earn return r_i, or defer earnings to earn more later, with proportional return $Y_i'(S)/Y_i(S)$

- Optimal schooling equalizes returns on these two investments

- Individual i's realized earnings are $Y_i(S_i^*)$

Ability Bias

- Empirical literature tries to estimate features of the potential earnings functions $Y_i(S)$
- Problem: As usual, we only see one earnings level for each person, corresponding to potential earnings at his/her chosen schooling level
- Why do people choose different levels of schooling? In the model differences must be driven either by variation in the discount rate, or in the potential earnings function
- "Ability bias:" Individuals that choose different schooling levels may have different potential earnings functions, leading observed returns to schooling to differ from causal returns
  - Label for selection bias in the returns to schooling context

Observed Returns to Schooling

- Consider an ordinary least squares (OLS) regression of observed earnings Y_i on schooling S_i:

$$Y_i = a + bS_i + e_i$$

- The observed return to schooling is the OLS slope coefficient:

$$b = \frac{Cov(Y_i, S_i)}{Var(S_i)}$$

- Question: Should I be worried about whether S_i is correlated with the error term e_i?

OLS Approximates the CEF

- Answer: No. By definition, the OLS residual e_i is orthogonal to the regressor S_i:

$$
\begin{aligned}
Cov(e_i, S_i) &= Cov(Y_i - a - bS_i, S_i) \\
&= Cov(Y_i, S_i) - bVar(S_i) \\
&= Cov(Y_i, S_i) - (Cov(Y_i, S_i)/Var(S_i))Var(S_i) \\
&= 0.
\end{aligned}
$$

- OLS always gives a minimum mean squared error approximation to the conditional expectation function (CEF), $E[Y_i|S_i]$:

$$(a, b) = \arg\min_{(a_0, b_0)} E\left[(E[Y_i|S_i] - a_0 - b_0 S_i)^2\right].$$

- OLS fits the CEF regardless of what model you have in mind. Better to ask: is the CEF economically interesting?

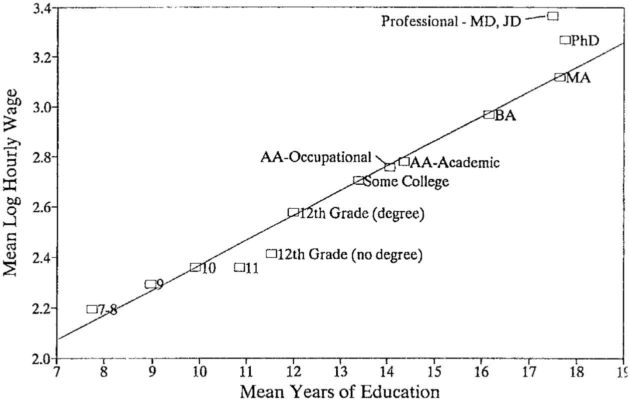


Fig. 2. Relationship between mean log hourly wages and completed education, men aged 40–45 in 1994–1996 Current Population Survey. Mean education by degree category estimated from February 1990 CPS.

Ability Bias

- Consider a constant effects potential earnings function:

$$Y_i(S) = \alpha_i + \beta S$$

- The causal return $\beta > 0$ is the same for all people and schooling levels
- This model implies

$$\frac{Y_i'(S_i^*)}{Y_i(S_i^*)} = r_i \implies S_i^* = \frac{1}{r_i} - \frac{\alpha_i}{\beta}$$

- Suppose the interest rate r_i is the same for everyone. Is the observed return to schooling too big or too small relative to the causal return?

Negative Ability Bias

- The observed return is too small
- When $r_i = r$ for all i, everyone earns the same amount:

$$Y_i(S_i^*) = \alpha_i + \beta\left(\frac{1}{r} - \frac{\alpha_i}{\beta}\right) = \frac{\beta}{r}.$$

- The observed return is therefore zero, which is less than the causal return β
- Intuition: The primary cost of schooling is the opportunity cost of earnings foregone. Higher-ability people face higher opportunity costs and so drop out earlier
- In this case "ability bias" is negative – the causal return exceeds the observed return

General Ability Bias

- More generally, the observed return to schooling is

$$b = \frac{Cov(Y_i(S_i^*), S_i^*)}{Var(S_i)}$$

$$= \frac{Cov\left(\frac{\beta}{r_i}, \frac{1}{r_i} - \frac{\alpha_i}{\beta}\right)}{Var\left(\frac{1}{r} - \frac{\alpha_i}{\beta}\right)}$$

$$= \beta \times \left(\frac{\sigma_{1/r}^2 - \sigma_{\alpha,1/r}/\beta}{\sigma_{1/r}^2 - 2\sigma_{\alpha,1/r}/\beta + \sigma_\alpha^2/\beta^2}\right).$$

- Ability bias depends on variances and covariances of discount rates and ability across people
- Direction is unclear *a priori*
- To get positive ability bias, need another force that overrides the basic opportunity cost story

Estimating Causal Returns

- The observed return to schooling may be contaminated by ability bias of unclear sign and magnitude. How can we estimate the causal return?
- Maintain the simple constant-effects model for potential earnings:

$$Y_i(S) = \alpha_i + \beta S$$

- We can then write observed earnings as

$$Y_i = \bar{\alpha} + \beta S_i + \epsilon_i.$$

- Here $\bar{\alpha} = E[\alpha_i]$ and $\epsilon_i = \alpha_i - \bar{\alpha}$
- Question: Should I be worried about whether S_i is correlated with the error term ϵ_i?

Observed and Causal Returns

$$Y_i = \bar{\alpha} + \beta S_i + \epsilon_i.$$

- Answer: Yes. The coefficient β is now defined as a parameter from a causal (potential outcomes) model, so there is no guarantee that $Cov(S_i, \epsilon_i) = 0$
- Schooling is not randomly assigned, so it may not be independent of potential outcomes, summarized here by ϵ_i
- This means the OLS slope coefficient b may not coincide with the causal effect β
- **Instrumental variables** (IV) is a common research design that seeks to eliminate selection bias in nonexperimental data

Instrumental Variables

$$Y_i = \bar{\alpha} + \beta S_i + \epsilon_i.$$

- Suppose we have a third variable, Z_i, that satisfies two conditions:
  1. **First stage:** $Cov(S_i, Z_i) \neq 0$.
  2. **Exclusion restriction:** $Cov(\epsilon_i, Z_i) = 0$.
- First stage requires Z_i (the instrument) to be correlated with S_i (the endogenous variable)
- Exclusion requires the instrument to be uncorrelated with potential outcomes (here, ϵ_i)
  - Z_i must be as good as randomly assigned
  - Z_i cannot affect Y_i through channels other than S_i

The Population IV Coefficient

- Covariance between outcome and instrument:

$$Cov(Y_i, Z_i) = Cov(\bar{\alpha} + \beta S_i + \epsilon_i, Z_i)$$

$$= \beta Cov(S_{i,} Z_i) + Cov(\epsilon_i, Z_i)$$

- Exclusion implies the second term is zero, so $Cov(Y_i, Z_i) = \beta Cov(S_i, Z_i)$
- First stage implies $Cov(S_i, Z_i) \neq 0$, so we can divide by this covariance to solve for β:

$$\frac{Cov(Y_i, Z_i)}{Cov(S_i, Z_i)} = \beta.$$

- The ratio of covariances on the left is the **population instrumental variables coefficient**, β_{IV}

IV Interpretation

- Divide the top and bottom of the IV coefficient by $Var(Z_i)$ to obtain:

$$\beta_{IV} = \frac{Cov(Y_i, Z_i)/Var(Z_i)}{Cov(S_i, Z_i)/Var(Z_i)}$$

- The IV coefficient is a ratio of two regression coefficients:
  - The **reduced form** regression of Y_i on Z_i
  - The **first stage** regression of S_i on Z_i
- Suppose Z_i is binary. Then the IV coefficient becomes a **Wald ratio** of two differences in means:

$$\beta_{IV} = \frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[S_i|Z_i = 1] - E[S_i|Z_i = 0]}.$$

IV Estimates of the Return to Schooling: Angrist and Krueger (1991)

- Angrist and Krueger (QJE 1991): classic study reporting IV estimates of the return to education
- Instrumental variables strategy motivated by interaction between compulsory schooling and age-at-entry laws
  - Students can typically drop out of school on the day they turn 16
  - Birth date cutoff for starting age: Students usually start school in the fall of the calendar year in which they turn six
- Generates differences in mean educational attainment by date of birth

Birth date	School start date	Dropout date	Schooling at dropout date
January 2, 1930	September 1, 1936	January 2, 1946	9.5 years
December 31, 1930	September 1, 1936	December 31, 1946	10.5 years

QOB Instruments

- AK's instrument is date of birth
- Operationalize using quarter of birth (QOB), which is available in US Census data
  - $Z_i = 1\{i \text{ was born in first quarter}\}$
- What do the first stage and exclusion restriction assumptions mean for a QOB instrument?

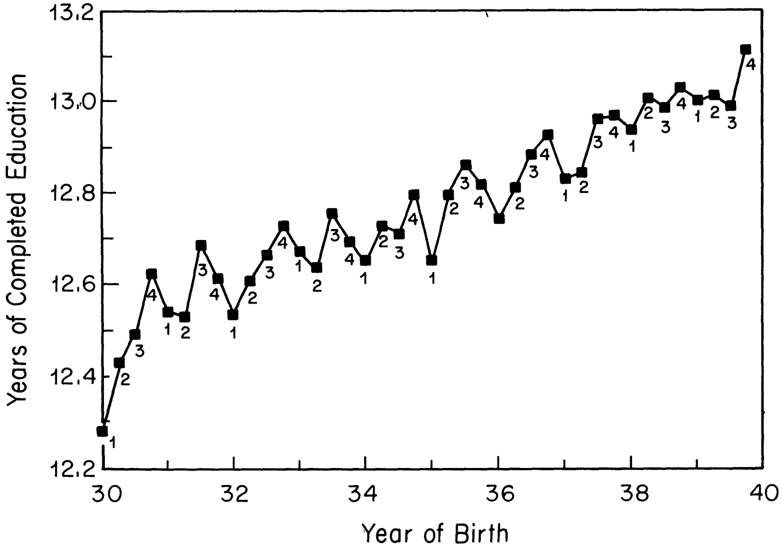


FIGURE I
Years of Education and Season of Birth
1980 Census
Note. Quarter of birth is listed below each observation.

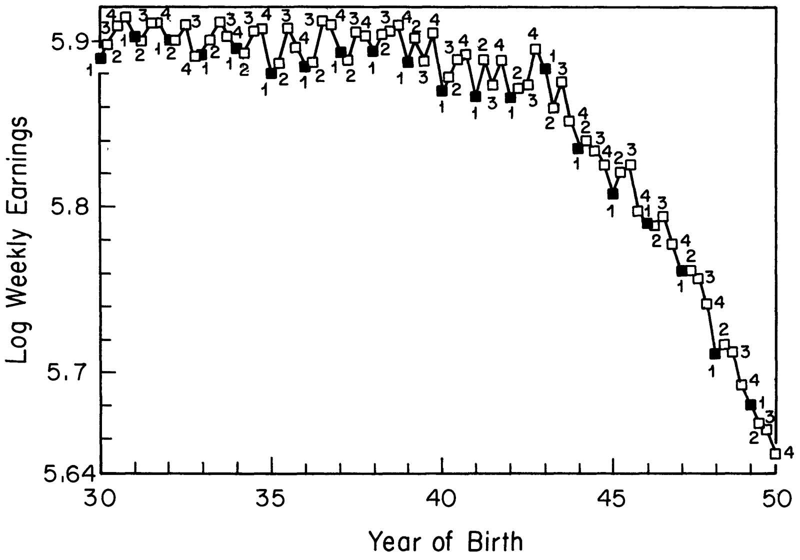


FIGURE V
Mean Log Weekly Wage, by Quarter of Birth
All Men Born 1930–1949; 1980 Census

TABLE III

PANEL A: WALD ESTIMATES FOR 1970 CENSUS—MEN BORN 1920–1929[a]

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.1484	5.1574	−0.00898
			(0.00301)
Education	11.3996	11.5252	−0.1256
			(0.0155)
Wald est. of return to education			0.0715
			(0.0219)
OLS return to education[b]			0.0801
			(0.0004)

Panel B: Wald Estimates for 1980 Census—Men Born 1930–1939

	(1) Born in 1st quarter of year	(2) Born in 2nd, 3rd, or 4th quarter of year	(3) Difference (std. error) (1) − (2)
ln (wkly. wage)	5.8916	5.9027	−0.01110
			(0.00274)
Education	12.6881	12.7969	−0.1088
			(0.0132)
Wald est. of return to education			0.1020
			(0.0239)
OLS return to education			0.0709
			(0.0003)

a. The sample size is 247,199 in Panel A, and 327,509 in Panel B. Each sample consists of males born in the United States who had positive earnings in the year preceding the survey. The 1980 Census sample is drawn from the 5 percent sample, and the 1970 Census sample is from the State, County, and Neighborhoods 1 percent samples.

b. The OLS return to education was estimated from a bivariate regression of log weekly earnings on years of education.

QOB Interpretation

- IV estimates based on QOB suggest a return to schooling of 7-10% per year
- IV estimates are comparable to or bigger than corresponding OLS estimates
- Card (1999) finds a similar pattern for other IV strategies
- In our simple model, this suggests negative ability bias: people with lower earnings potential attend school for longer
- Other interpretations?

Heterogeneous Treatment Effects

- Our simple model assumed constant effects of schooling across people
- Return to general potential outcomes model with binary treatment D_i and potential outcomes $Y_i(1)$ and $Y_i(0)$
- Suppose we have a binary instrument Z_i, and consider two new potential outcomes defined by a hypothetical manipulation of Z_i:
  - $D_i(1)$: i's treatment status if $Z_i = 1$
  - $D_i(0)$: i's treatment status if $Z_i = 0$
- Observed treatment is $D_i = D_i(0) + (D_i(1) - D_i(0))Z_i$

IV Assumptions

- IV assumptions in a heterogeneous treatment effects world:
    1. **Independence/exclusion:** $(Y_i(1), Y_i(0), D_i(1), D_i(0)) \perp\!\!\!\perp Z_i$
    2. **First stage:** $\Pr[D_i = 1 | Z_i = 1] > \Pr[D_i = 1 | Z_i = 0]$
    3. **Monotonicity:** $D_i(1) \geq D_i(0)\ \forall i$
- Relative to our constant effects IV setup, monotonicity is the novel assumption
- Monotonicity requires the instrument to affect everyone's treatment status in the same direction

Compliance Groups

- Under monotonicity, we can partition the population into three groups defined by their behavioral responses to the instrument (Angrist, Imbens, and Rubin 1996):

  1. **Always takers:** $D_i(1) = D_i(0) = 1$
  2. **Never takers**: $D_i(1) = D_i(0) = 0$
  3. **Compliers:** $D_i(1) = 1$, $D_i(0) = 0$

- Compliers have $D_i(1) > D_i(0)$: their treatment status increases with the instrument

- Monotonicity rules out **defiers** with $D_i(1) = 0$, $D_i(0) = 1$

Local Average Treatment Effects

- Under these assumptions, IV identifies a **local average treatment effect** (LATE):

$$\frac{E\left[Y_i|Z_i=1\right]-E[Y_i|Z_i=0]}{E[D_i|Z_i=1]-E[D_i|Z_i=0]}=E[Y_i(1)-Y_i(0)|D_i(1)>D_i(0)]$$

- This is the LATE theorem of Imbens and Angrist (1994)
- LATE is the average treatment effect for compliers – individuals whose treatment status is determined by the instrument

LATE Theorem: Proof

- Note that $Y_i = Y_i(D_i) = Y_i(D_i(Z_i))$, so by independence

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(D_i(1))|Z_i = 1] - E[Y_i(D_i(0))|Z_i = 0]$$

$$= E[Y_i(D_i(1)) - Y_i(D_i(0))].$$

- By monotonicity we either have $D_i(1) = D_i(0)$ or $D_i(1) > D_i(0)$, so

$$E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0] = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)]\Pr[D_i(1) > D_i(0)]$$

- The same logic implies $E[D_i|Z_i = 1] - E[D_i|Z_i = 0] = \Pr[D_i(1) > D_i(0)]$, so

$$\frac{E[Y_i|Z_i = 1] - E[Y_i|Z_i = 0]}{E[D_i|Z_i = 1] - E[D_i|Z_i = 0]} = E[Y_i(1) - Y_i(0)|D_i(1) > D_i(0)].$$

Interpreting IV Estimates

- LATE interpretation suggests that QOB instrument identifies the causal effect of extra schooling for individuals on the margin of dropping out early around mid-century
- Next, we will consider more recent evidence looking at other schooling margins

Returns to College for Marginal Students: Zimmerman (2014)

- Observed return to college has increased dramatically in recent decades
  - College wage premium rose from 50% to 97% between 1980 and 2008 (Acemoglu and Autor, 2011)
  - May reflect fast growth of skill demand coupled with slow growth of skill supply (Goldin and Katz, 2008)
- At the same time, many students in the US start college but don't finish
  - 62% of students attending four-year colleges graduate within 6 years (NCES, 2020)
  - Does college attendance improve earnings for academically marginal students?
- Zimmerman (JOLE 2014) leverages a **regression discontinuity design** to study returns for students on the margin of four-year college enrollment

Regression Discontinuity Designs

- Consider a setting with a binary treatment $D_i \in \{0, 1\}$, and potential outcomes $Y_i(1)$ and $Y_i(0)$
- Suppose the treatment is a deterministic and discontinuous function of an observed covariate R_i, such that

$$D_i = 1\{R_i > c\}.$$

- R_i is called the **running variable** or **forcing variable**
- This is a **sharp RD** because the probability of treatment switches from zero to one at the threshold
- Zimmerman (2014): GPA cutoff for admission to state universities in Florida

Regression Discontinuity Designs

- We get to observe $Y_i(1)$ when $R_i > c$ and $Y_i(0)$ when $R_i \leq c$
- Basic idea of the RD design: Compare observations just above and just below the threshold to infer treatment effect
- Intuitively, the treatment may be as good as randomly assigned for individuals in the neighborhood of $R_i = c$, so comparing treated and nontreated near c reveals a treatment effect

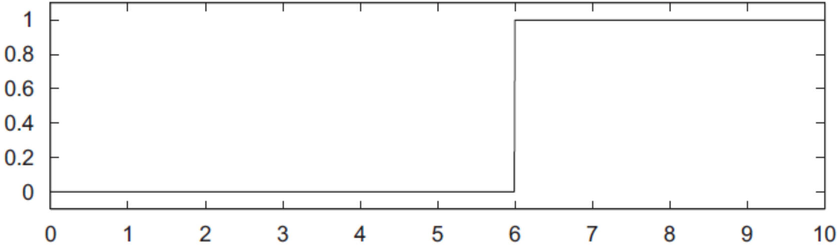


Fig. 1. Assignment probabilities (SRD).

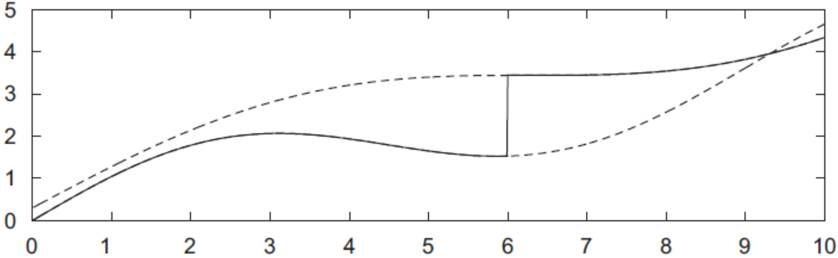


Fig. 2. Potential and observed outcome regression functions.

Source: Imbens and Lemieux (2008)

RD Identification

- Key assumption: potential outcomes are smooth at the threshold
- Formally:

$$\lim_{r \to c^+} E\left[Y_i(d)|R_i = r\right] = \lim_{r \to c^-} E\left[Y_i(d)|R_i = r\right],\ d \in \{0, 1\}$$

- Potential outcome CEFs must be continuous at the threshold
- The population just below must not be discretely different from the population just above

RD Identification

- If this assumption holds we have

$$\lim_{r \to c^+} E\left[Y_i | R_i = r\right] - \lim_{r \to c^-} E\left[Y_i | R_i = r\right]$$

$$= \lim_{r \to c^+} E\left[Y_i(1) | R_i = r\right] - \lim_{r \to c^-} E\left[Y_i(0) | R_i = r\right]$$

$$= E\left[Y_i(1) | R_i = c\right] - E\left[Y_i(0) | R_i = c\right]$$

$$= E\left[Y_i(1) - Y_i(0) | R_i = c\right]$$

- When potential outcomes are smooth around the threshold, a comparison of individuals just above and just below yields the average treatment effect for those at the threshold
- Identification argument is nonparametric: we don't need to assume anything about the distribution of potential outcomes other than continuity of CEFs

RD Interpretation

- Core RD intuition: for those near the threshold, things could have gone either way
- Interpret RD as a local randomized trial among those sufficiently close to $R_i = c$
- Explains why RD evidence can be especially compelling relative to other research designs – close to RCT ideal
- "Local randomization" view motivates common RD diagnostics
  - Check balance of pre-determined characteristics for observations above and below the threshold
  - Look for anomalies in the distribution of the running variable around the threshold, which may indicate manipulation (McCrary, 2008)

Fuzzy RD

- Sometimes treatment is generated by a discontinuous assignment rule that isn't deterministic
- Suppose that

$$\lim_{r \to c^-} Pr\left[D_i = 1 | R_i = r\right] < \lim_{r \to c^+} Pr\left[D_i = 1 | R_i = r\right]$$

- The probability of treatment jumps at $R_i = c$, but not necessarily from zero to one
- This is a **fuzzy RD** scenario because treatment is only partly determined by the threshold
- Zimmerman (2014): Students above GPA cutoff are eligible for admission, but not guaranteed

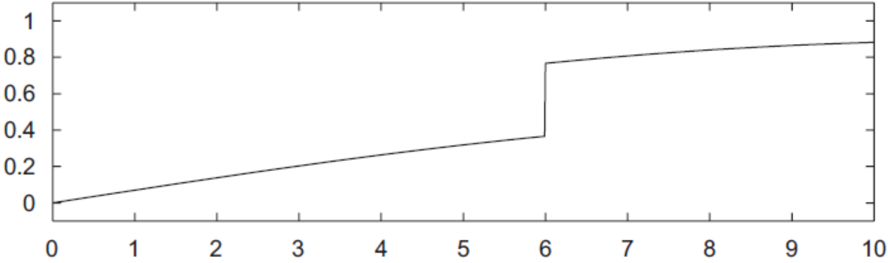


Fig. 3. Assignment probabilities (FRD).

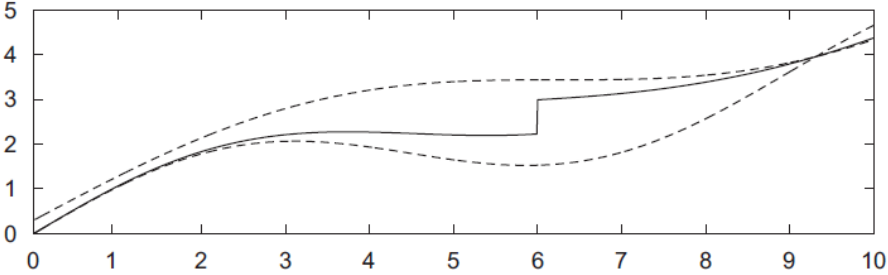


Fig. 4. Potential and observed outcome regression (FRD).

Fuzzy RD Assumptions

- As before, assume the distributions of $Y_i(1)$ and $Y_i(0)$ are smooth around the threshold
- Let $D_i(1)$ and $D_i(0)$ denote potential treatment statuses for individual i if s/he were located above and below the threshold. Assume these are also smooth across the threshold, and

$$D_i(1) \geq D_i(0) \ \forall i$$

- Crossing the threshold weakly increases the likelihood of treatment for everyone

Fuzzy RD

- Under these assumptions, we have

$$\frac{\lim_{r \to c^+} E\left[Y_i|R_i = r\right] - \lim_{r \to c^-} E\left[Y_i|R_i = r\right]}{\lim_{r \to c^+} E\left[D_i|R_i = r\right] - \lim_{r \to c^-} E\left[D_i|R_i = r\right]}$$

$$= E\left[Y_i(1) - Y_i(0)|D_i(1) > D_i(0), R_i = c\right]$$

- The numerator on the left is the jump in outcomes at the threshold, as in a sharp RD
- The denominator is the change in the probability of treatment at the threshold
- The ratio of the jump in the outcome CEF to the jump in the treatment probability identifies an average treatment effect for individuals who switch treatment status at the threshold
- Sound familiar?

Fuzzy RD is IV

- Fuzzy RD is IV using a threshold indicator $Z_i = 1\{R_i > c\}$ as an instrument for treatment in the neighborhood of the threshold
- Think of fuzzy RD as a local randomized trial with non-compliance
- Implies fuzzy RD estimates are local in two senses
  - Local to the threshold, $R_i = c$ (also applies to sharp RD)
  - Only apply to compliers at the threshold (that's the "local" in LATE)

RD Implementation

- Implementing RD requires estimating the left- and right-hand limits of average outcomes and treatment probabilities
- Bias/variance tradeoff: using data away from the threshold increases sample size, but may introduce bias if potential outcomes are related to the running variable
- In practice RD is usually implemented with local linear regression
  - Regress outcome on the running variable among observations within a small bandwidth of the threshold, with weights that decline with distance to threshold
  - RD estimate is difference in fitted regression functions above and below the threshold
- Recent econometric literature proposes optimal bandwidths that balance bias and variance to minimize mean squared error, automated in *rdrobust* Stata command (Imbens and Kalyanaraman, 2011; Calonico et al., 2014)

Returns to College for Marginal Students: Zimmerman (2014)

- Zimmerman (2014) uses a GPA cutoff to estimate the returns to four-year college admission at public institutions in Florida
- Students above the cutoff are eligible for admission to schools in the Florida State University System (SUS)
- In practice, the cutoff is relevant for admission to Florida International University (FIU), a large SUS campus in Miami
- Population around the FIU admission cutoff has relatively low SAT scores (21st percentile nationwide) and low graduation rates
- Estimates are therefore informative about returns to college for marginal students

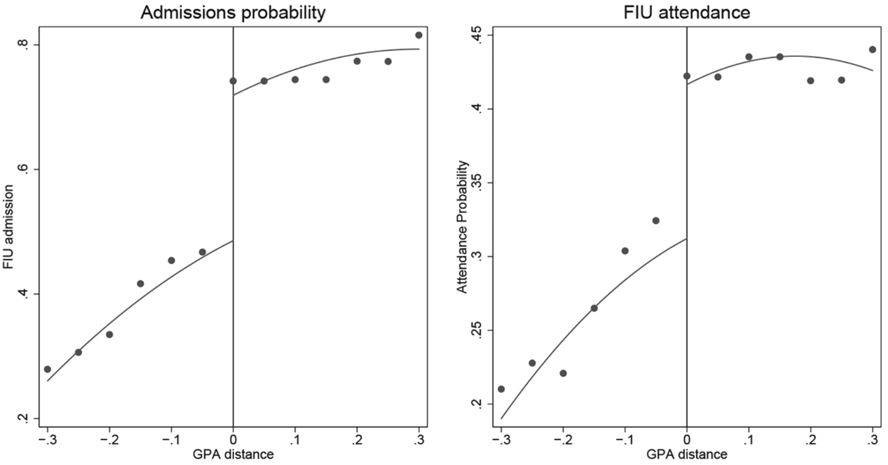


FIG. 4.—Admissions and FIU attendance. Lines are fitted values based on the main specification. Dots, shown every .05 grade points, are rolling averages of values within .05 grade points on either side that have the same value of the threshold-crossing dummy.

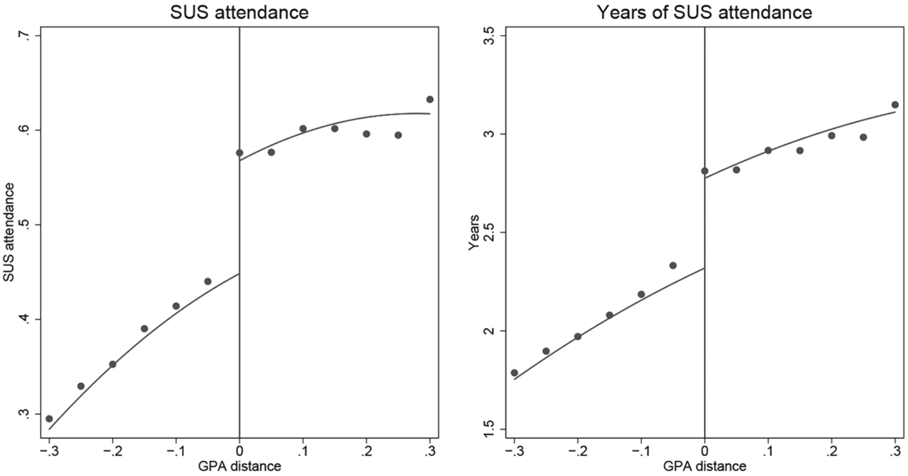


FIG. 5.—SUS attendance and persistence. Lines are fitted values based on the main specification. Dots, shown every .05 grade points, are rolling averages of values within .05 grade points on either side that have the same value of the threshold-crossing dummy.

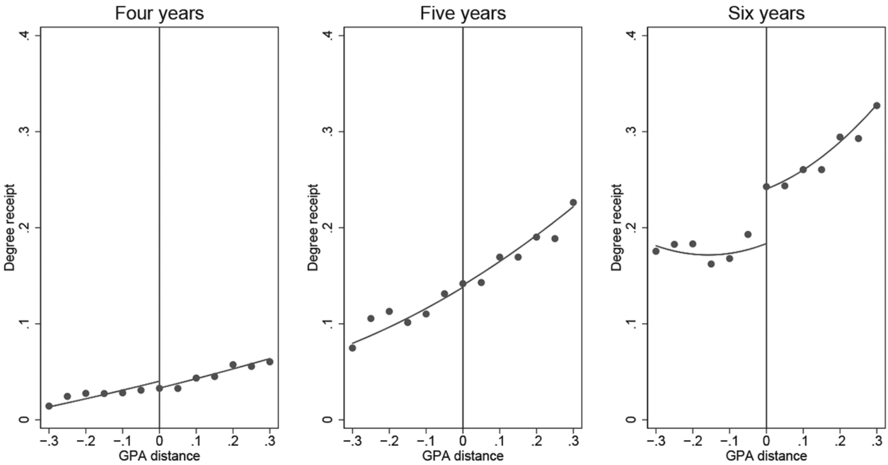


FIG. 6.—SUS BA receipt by years elapsed since high school. Lines are fitted values based on the main specification. Dots, shown every .05 grade points, are rolling averages of values within .05 grade points on either side that have the same value of the threshold-crossing dummy.

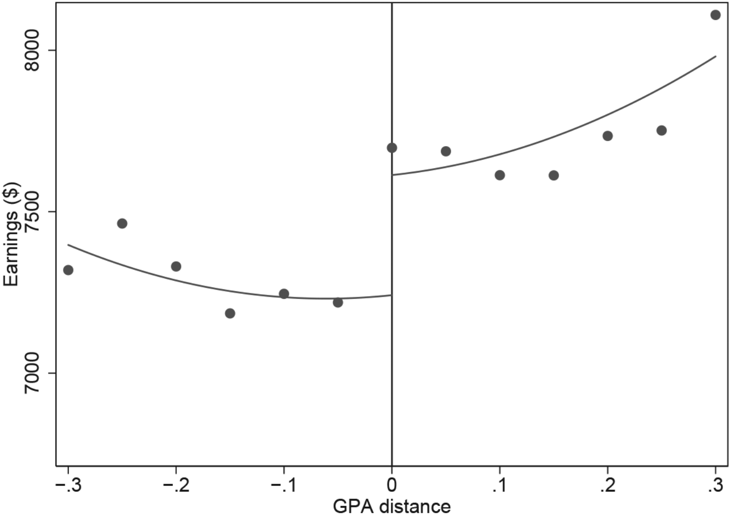


FIG. 8.—Quarterly earnings by distance from GPA cutoff. Lines are fitted values based on the main specification. Dots, shown every .05 grade points, are rolling averages of values within .05 grade points on either side that have the same value of the threshold-crossing dummy.

Table 5
Earnings Effects 8–14 Years after High School Completion

	Main	Controls	BW=.5	BW=.15	Local Linear
Reduced-form estimates:					
Above cutoff	372*	366**	409**	479**	410**
	(141)	(130)	(154)	(198)	(147)
Instrumental variables estimates:					
FIU admission	1,593*	1,575**	1,665**	1,700**	2,001*
	(604)	(584)	(645)	(621)	(696)
Years of SUS attendance	815**	792**	833**	966***	977**
	(276)	(262)	(271)	(305)	(306)
BA degree	6,547*	6,442*	7,366*	10,769	5,958**
	(2,496)	(2,411)	(2,998)	(5,726)	(2,024)
N	6,542	6,542	9,659	3,294	6,542

NOTE.—FIU = Florida International University; SUS = State University System; BA = bachelor's degree. Standard errors are clustered within grade bins. The *p*-values are calculated using a clustered wild bootstrap-*t* procedure described in Sec. III and app. B. The dependent variable in each regression is average quarterly earmings in 2005 dollars. The "BW=.15" specification uses observations within .15 grade points above and below the cutoff and allows for a linear trend in distance from the cutoff. The "BW=.5" specification uses observations within the .5 grade points on either side of the cutoff and allows for a quartic polynomial in distance from the cutoff. The "Local Linear" specification is identical to the main specification, but it allows for linear slope terms in distance from the cutoff that differ above and below the threshold.

* Significant at the 10% level.
** Significant at the 5% level.
*** Significant at the 1% level.

Human Capital vs. Signaling

- Evidence so far suggests that education increases earnings
- Conventional human capital view is that schooling investments raise earnings by boosting productivity
- **Signaling models** (Spence, 1973) provide an alternative explanation for the return to schooling
  - If employers cannot observe ability, schooling may serve as a costly signal that separates low- and high-ability types, rather than increasing productivity
  - Implies schooling is pure social waste: burns resources to create inequality
- Distinguishing between human capital and signaling views is essential for education policy
- Signaling models provide an explanation for "sheepskin effects:" observed return to schooling is especially large for grade 12

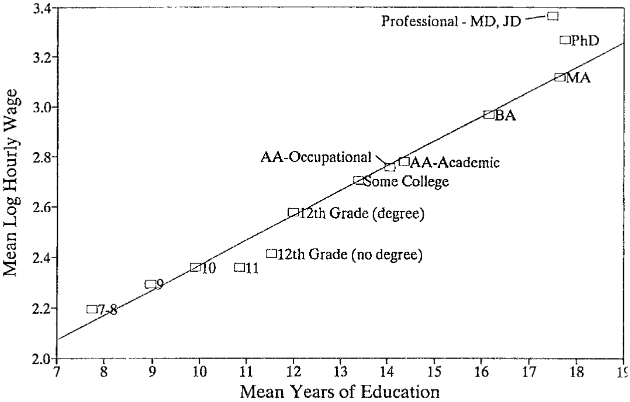


Fig. 2. Relationship between mean log hourly wages and completed education, men aged 40–45 in 1994–1996 Current Population Survey. Mean education by degree category estimated from February 1990 CPS.

Signaling Value of a High School Diploma: Clark and Martorell (2014)

- Clark and Martorell (JPE 2014) use an RD design to estimate the causal effect of high school graduation on earnings
- CM use the fact that students in Texas must pass exams before graduating high school
- Testing starts in 10th grade and students can try multiple times, but eventually face a "last chance" exam at the end of 12th grade
- Students who just barely fail vs. barely pass should have similar human capital, but differ in educational credentials
- RD therefore plausibly identifies the signaling value of a diploma
- There is some "slippage" even with last-chance exams – so the RD is fuzzy

TABLE 6

Associations between Diploma and Test Scores and Earnings

	A. Mean Differences by Diploma Status				
	Last-Chance Sample	Complete Grade 12, No College			
		All	T1	T2	T3
	(1)	(2)	(3)	(4)	(5)
Earnings years 7–11	1,814.7	2,867.8	1,780.3	1,752.0	2,385.3
	(138.1)	(79.3)	(111.8)	(176.1)	(228.5)
Observations	128,460	992,031	210,793	193,970	194,896
Mean earnings without diploma	12,400	12,673	11,858	13,301	13,538
Difference (%)	14.6	22.6	15.0	13.2	17.6
PDV earnings	8,054.5	8,731.0	7,280.7	7,459.4	10,546.3
	(632.3)	(341.9)	(501.9)	(779.8)	(951.4)
Observations	37,571	340,028	74,490	63,652	64,548
Mean earnings without diploma	70,280	69,992	66,466	74,216	73,860
Difference (%)	11.5	12.5	11.0	10.1	14.3

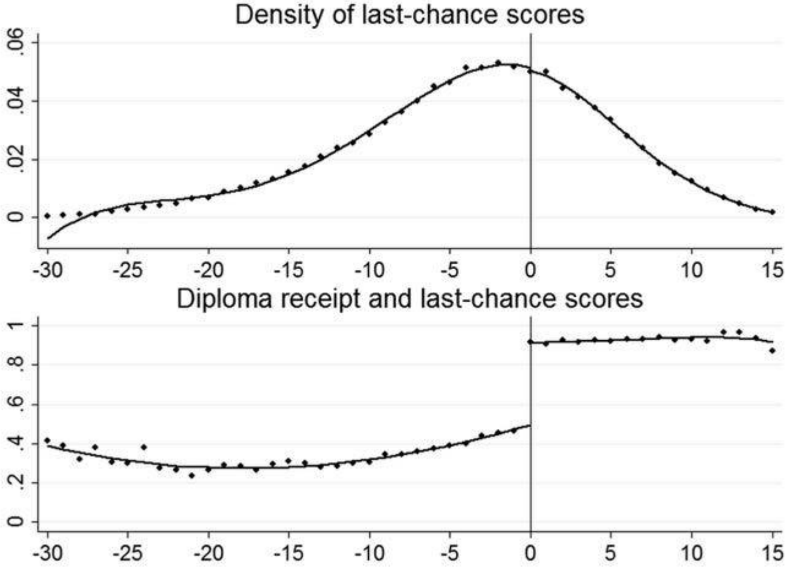
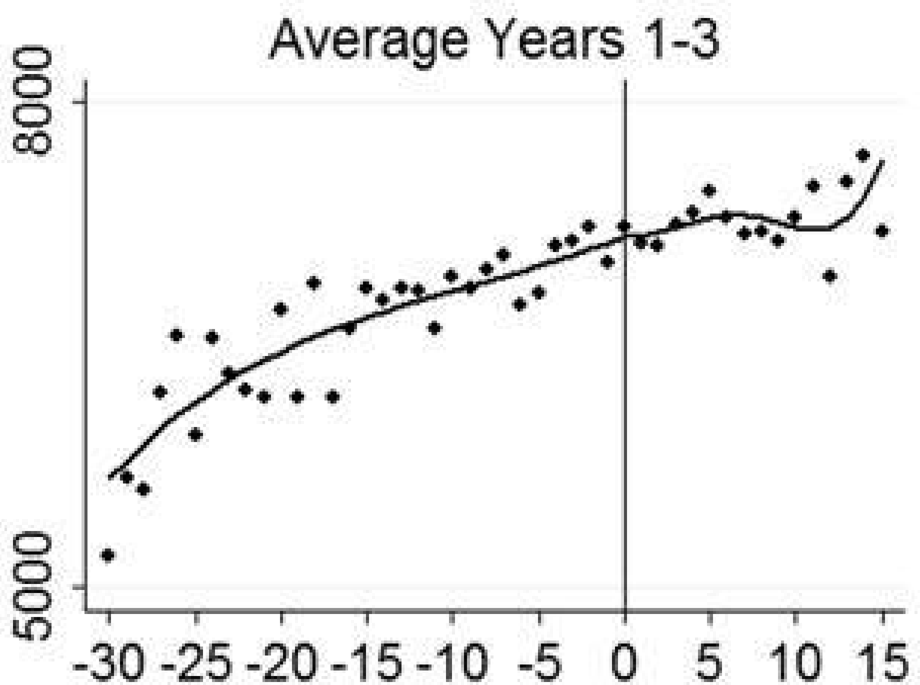
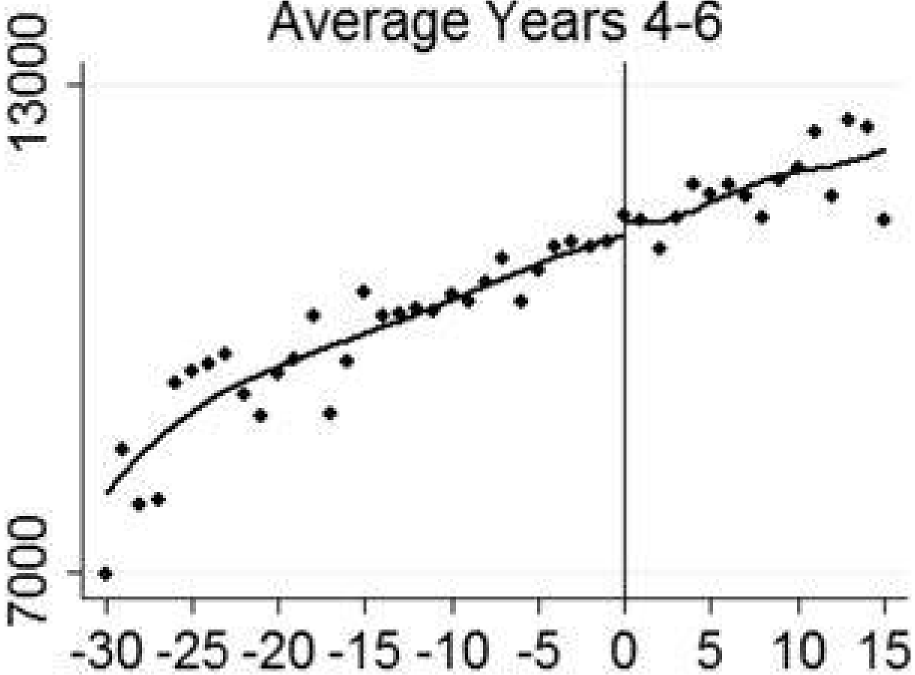
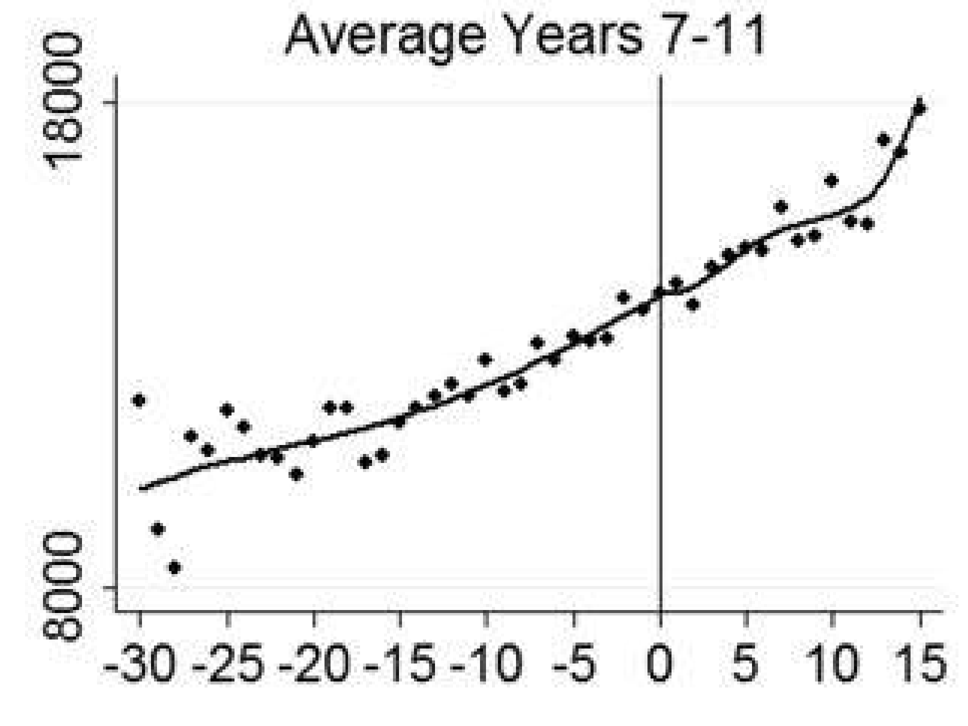


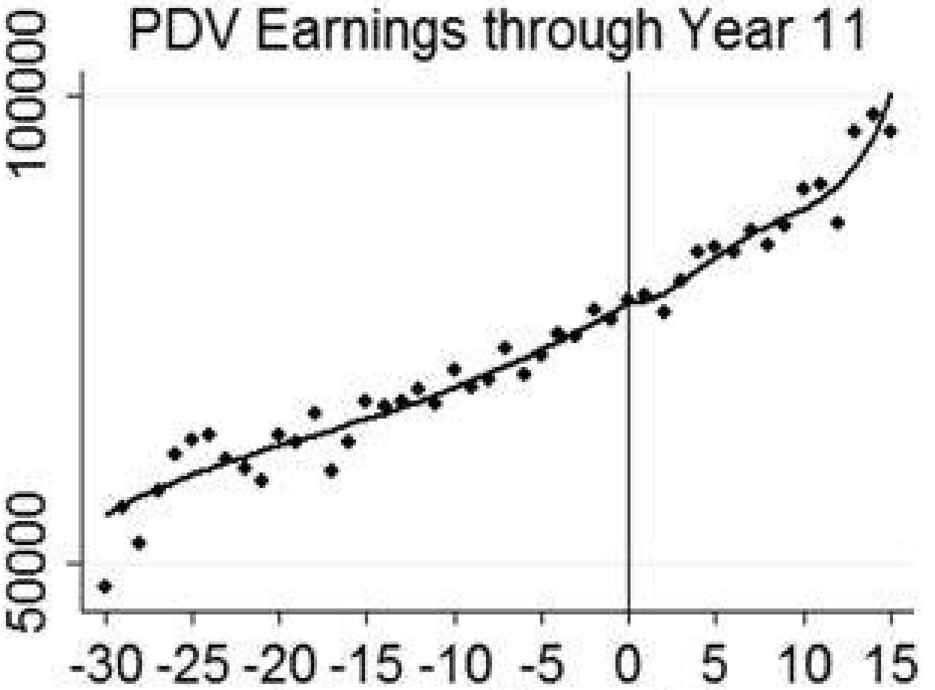
FIG. 1.—Last-chance exam scores and diploma receipt. The graphs are based on the last-chance sample. See table 1 and the text. Dots are test score cell means. The scores on the x-axis are the minimum of the section scores (recentered to be zero at the passing cutoff) that are taken in the last-chance exam. Lines are fourth-order polynomials fitted separately on either side of the passing threshold.



Average Years 1-3
8000
5000
-30
-25
-20
-15
-10
-5
0
5
10
15

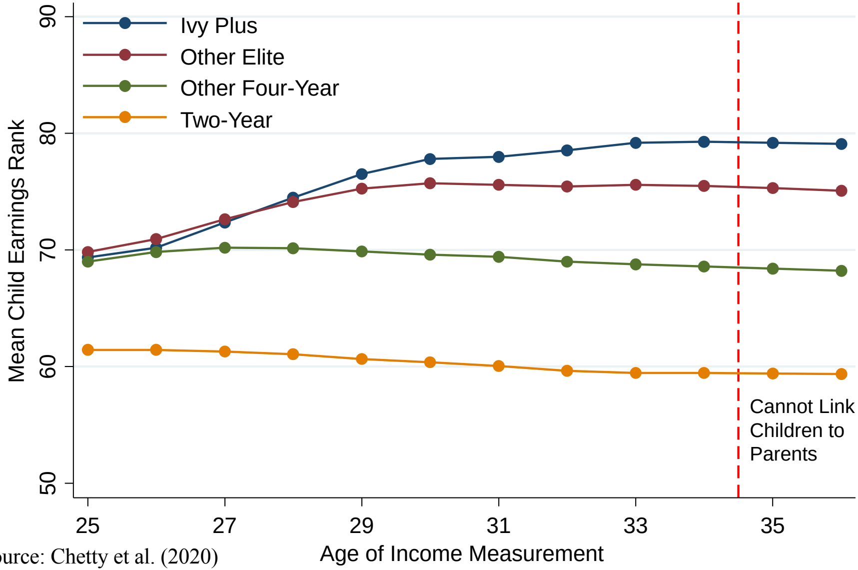

Average Years 4-6
13000
7000
-30 -25 -20 -15 -10 -5 0 5 10 15


Average Years 7-11
18000
8000
-30
-25
-20
-15
-10
-5
0
5
10
15

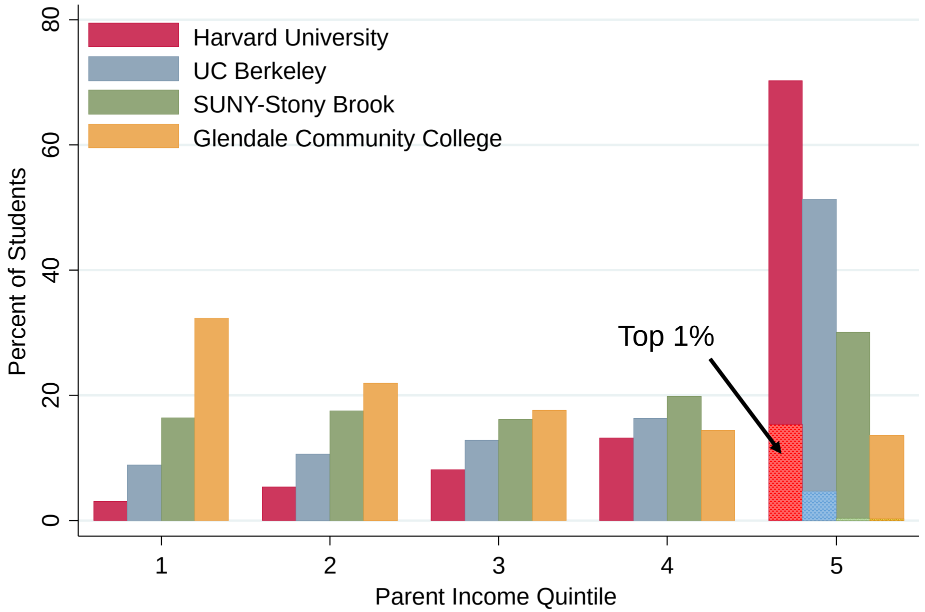

PDV Earnings through Year 11
100000
50000
-30
-25
-20
-15
-10
-5
0
5
10
15

Returns to College Selectivity

- For many students the relevant choice margin is which college to attend rather than years of schooling or college vs. no college
- Very large differences in earnings between students attending different US colleges
- But there is also a lot of selection into college choice



Source: Chetty et al. (2020)


Harvard University
UC Berkeley
SUNY-Stony Brook
Glendale Community College
Top 1%
Percent of Students
0
20
40
60
80
1
2
3
4
5
Parent Income Quintile

Returns to College Selectivity

- Hard to find good experiments and quasi-experiments that induce variation in attendance at more vs. less selective colleges
- Dale and Krueger (2002, 2014) use a **matching** approach that compares outcomes for students who applied and were admitted to the same sets of colleges, but attended different schools
- Based on a **selection on observables** assumption: college choice is independent of potential outcomes conditional on a set of observed covariates

Potential Outcomes Model

- Return to our causal model with binary treatment $D_i \in \{0, 1\}$ and potential outcomes $Y_i(1)$ and $Y_i(0)$
- Suppose treatment isn't randomly assigned
- As we've seen, the observed difference between average outcomes for individuals with $D_i = 1$ and $D_i = 0$ may be contaminated by selection bias
- Suppose we also have data on a vector of observed covariates X_i
- Dale and Krueger: D_i is attending a more selective college, and X_i is the lists of colleges where a student applied and was admitted

Selection on Observables

- Selection-on-observables approaches are based on a **conditional independence assumption** (CIA):

$$(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i | X_i$$

- CIA is also called "unconfoundedness," "ignorability," "exogeneity"
- The idea is that while potential outcomes and treatment may not be independent in general, they are independent conditional on a set of observed covariates - treatment is as good as random conditional on X_i
- CIA necessarily holds in stratified RCTs, and may hold in non-experimental data with the right controls

Full Covariate Matching

- Under CIA an obvious approach is to simply compare treatment and control groups conditional on the covariates
- Let $\Delta(x)$ denote the observed treatment/control difference for a particular value of the covariates:

$$\Delta(x) \equiv E\left[Y_i | D_i = 1, X_i = x\right] - E\left[Y_i | D_i = 0, X_i = x\right]$$

- CIA implies

$$\begin{aligned}\Delta(x) &= E\left[Y_i(1) | D_i = 1, X_i = x\right] - E\left[Y_i(0) | D_i = 0, X_i = x\right] \\ &= E\left[Y_i(1) - Y_i(0) | X_i = x\right] \\ &\equiv ATE(x).\end{aligned}$$

- Covariate-specific treatment/control contrasts capture conditional average treatment effects
- By computing $\Delta(x)$ for every value of x and then weighting appropriately, we can obtain any causal effect of interest. This is **full covariate matching**

Computing Treatment Effects

- Under CIA, we can use full covariate matching to compute average treatment effects:

$$ATE = \sum_x Pr\left[X_i = x\right] \Delta(x)$$

$$TOT = \sum_x Pr\left[X_i = x | D_i = 1\right] \Delta(x)$$

$$TNT = \sum_x Pr\left[X_i = x | D_i = 0\right] \Delta(x)$$

OLS Regression as Matching

- Consider an OLS regression of outcomes on a treatment indicator, controlling for indicators for every value of the covariates X_i:

$$Y_i = a + bD_i + \sum_x \pi_x 1\{X_i = x\} + e_i$$

- This regression is **saturated** in the controls: there is a different coefficient for every value of X_i

- With saturated controls, the OLS coefficient is

$$b = \sum_x \left(\frac{\Pr[X_i=x] Var(S_i|X_i=x)}{\sum_{x'} \Pr[X_i=x'] Var(S_i|X_i=x')} \right) \Delta(x).$$

- OLS with saturated controls is a version of full covariate matching
  - "Saturate-and-weight" theorem (Angrist and Pischke, 2009)
  - Under CIA, generates a variance-weighted average treatment effect

CIA Methods

- In practice, full covariate matching may not be feasible (e.g. many-valued or continuous controls)
- There are a variety of approaches to controlling for X_i in such cases:
  - OLS with additive controls
  - Nearest-neighbor or kernel matching
  - Propensity score matching/reweighting
- These methods are not qualitatively different
  - All are approaches to adjusting for covariates
  - Coincide when the controls are flexible enough
- Key to the research design is the underlying CIA assumption, not the particular method used to control for X_i

Returns to College Selectivity: Dale and Krueger

- Dale and Krueger (QJE 2002, JHR 2014) take a matching/selection on observables approach to estimating the returns to college selectivity
- Research design: compare students who applied to, and were admitted by, the same colleges, but chose to attend different schools
- Intuition: Application choices capture a lot of students' information about their own ability, while admission decisions capture a lot of colleges' information about student ability
- Data: College and Beyond (C&B)
  - Survey of students enrolled at 34 colleges, more selective than the US average
  - 2014 paper matches C&B to administrative earnings data from the Social Security Administration (SSA)

TABLE I

ILLUSTRATION OF HOW MATCHED-APPLICANT GROUPS WERE CONSTRUCTED

		Student applications to college							
		Application 1		Application 2		Application 3		Application 4	
Student	Matched-applicant group	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision	School average SAT	School admissions decision
Student A	1	1280	Reject	1226	Accept*	1215	Accept	na	na
Student B	1	1280	Reject	1226	Accept	1215	Accept*	na	na
Student C	2	1360	Accept	1310	Reject	1270	Accept*	1155	Accept
Student D	2	1355	Accept	1316	Reject	1270	Accept*	1160	Accept
Student E	2	1370	Accept*	1316	Reject	1260	Accept	1150	Accept
Student F	Excluded	1180	Accept*	na	na	na	na	na	na
Student G	Excluded	1180	Accept*	na	na	na	na	na	na
Student H	3	1360	Accept	1308	Accept*	1260	Accept	1160	Accept
Student I	3	1370	Accept*	1311	Accept	1255	Accept	1155	Accept
Student J	3	1350	Accept	1316	Accept*	1265	Accept	1155	Accept
Student K	4	1245	Reject	1217	Reject	1180	Accept*	na	na
Student L	4	1235	Reject	1209	Reject	1180	Accept*	na	na
Student M	5	1140	Accept	1055	Accept*	na	na	na	na
Student N	5	1145	Accept*	1060	Accept	na	na	na	na
Student O	No match	1370	Reject	1038	Accept*	na	na	na	na

* Denotes school attended.

na = did not report submitting application.

The data shown on this table represent hypothetical students. Students F and G would be excluded from the matched-applicant subsample because they applied to only one school (the school they attended). Student O would be excluded because no other student applied to an equivalent set of institutions.

TABLE V

LINEAR REGRESSIONS PREDICTING WHETHER STUDENT ATTENDED MOST SELECTIVE COLLEGE FOR C&B SAMPLE OF STUDENTS ADMITTED TO MORE THAN ONE SCHOOL

	Parameter estimates	
	Matched-applicant model*	Self-revelation model
Predicted log (parental income)	−0.024	−0.037
	(0.026)	(0.030)
Own SAT score/100	0.020	0.021
	(0.005)	(0.007)
Female	0.034	0.033
	(0.014)	(0.028)
Black	0.056	−0.005
	(0.026)	(0.037)
Hispanic	−0.019	0.042
	(0.064)	(0.074)
Asian	0.019	0.074
	(0.026)	(0.050)
Other/missing race	−0.095	0.010
	(0.093)	(0.081)
High school top 10 percent	−0.014	−0.020
	(0.021)	(0.028)
High school rank missing	−0.035	−0.040
	(0.036)	(0.058)
Athlete	0.056	0.059
	(0.023)	(0.045)
Average SAT score/100 of schools		−0.122
applied to		(0.040)
One additional application		0.149
		(0.037)
Two additional applications		0.076
		(0.033)
Three additional applications		0.020
		(0.038)
N	5536	8257

TABLE III

LOG EARNINGS REGRESSIONS USING COLLEGE AND BEYOND SURVEY, SAMPLE OF MALE AND FEMALE FULL-TIME WORKERS

	Model					
	Basic model: no selection controls		Matched-applicant model	Alternative matched-applicant models		Self-revelation model
	Full sample	Restricted sample	Similar school-SAT matches*	Exact school-SAT matches**	*Barron's* matches***	
Variable	1	2	3	4	5	6
School-average SAT score/100	0.076	0.082	−0.016	−0.106	0.004	−0.001
	(0.016)	(0.014)	(0.022)	(0.036)	(0.016)	(0.018)
Predicted log(parental income)	0.187	0.190	0.163	0.232	0.154	0.161
	(0.024)	(0.033)	(0.033)	(0.079)	(0.028)	(0.025)
Own SAT score/100	0.018	0.006	−0.011	0.003	−0.005	0.009
	(0.006)	(0.007)	(0.007)	(0.014)	(0.005)	(0.006)
Female	−0.403	−0.410	−0.395	−0.476	−0.400	−0.396
	(0.015)	(0.018)	(0.024)	(0.049)	(0.017)	(0.014)
Black	−0.023	−0.026	−0.057	−0.028	−0.057	−0.034
	(0.035)	(0.053)	(0.053)	(0.049)	(0.039)	(0.035)
Hispanic	0.015	0.070	0.020	−0.248	0.036	0.007
	(0.052)	(0.076)	(0.099)	(0.206)	(0.066)	(0.053)
Asian	0.173	0.245	0.241	0.368	0.163	0.155
	(0.036)	(0.054)	(0.064)	(0.141)	(0.049)	(0.037)
Other/missing race	−0.188	−0.048	0.060	−0.072	−0.050	−0.192
	(0.119)	(0.143)	(0.180)	(0.083)	(0.134)	(0.116)
High school top 10 percent	0.061	0.091	0.079	0.091	0.079	0.063
	(0.018)	(0.022)	(0.026)	(0.032)	(0.024)	(0.019)
High school rank missing	0.001	0.040	0.016	0.029	0.025	−0.009
	(0.024)	(0.026)	(0.038)	(0.066)	(0.027)	(0.022)
Athlete	0.102	0.088	0.104	0.169	0.093	0.094
	(0.025)	(0.030)	(0.039)	(0.096)	(0.033)	(0.024)
Average SAT score/100 of schools applied to						0.090
						(0.013)
One additional application						0.064
						(0.011)
Two additional applications						0.074
						(0.022)
Three additional applications						0.112
						(0.028)
Four additional applications						0.085
						(0.027)
Adjusted R^2	0.107	0.110	0.112	0.142	0.106	0.113
N	14,238	6,335	6,335	2,330	9,202	14,238

Table 3

Comparing Parameter Estimates of the Effect of College Average SAT Score on Earnings Using C&B and SSA Data, 1976 Cohort

	C&B sample[a]		Merged C&B and SSA sample[b]									
	Log 1995 C&B earnings		Log 1995 C&B earnings		Log 1995 SSA earnings (topcoded)		Log 1995 SSA earnings (not topcoded)		Log (median of 1993 to 1997 earnings), SSA data		Log (median of 1993 to 1997 earnings), SSA data	
	1	2	3	4	5	6	7	8	9	10	11	12
	Basic	Self – revelation	Basic	Self - revelation	Basic	Self – revelation	Basic	Self - revelation	Basic	Self – revelation	Basic	Self – revelation
Parameter estimate for school SAT/100	0.076	-0.001	0.068	-0.007	0.048	-0.021	0.058	-0.015	0.059	-0.025	0.061	-0.023
	(.008)	(.012)	(.007)	(.012)	(.009)	(.014)	(.009)	(.015)	(.008)	(.012)	(.007)	(.012)
	{.016}	{.018}	{.014}	{.018}	{.016}	{.018}	{.017}	{.016}	{.012}	{.013}	{.013}	{.014}
N	14,238		10,886		10,886		10,886		11,932		12,075	
Sample restriction	Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Full-time workers (according to C&B survey)		Median earnings greater than zero (SSA data)		Median earnings greater than $13,822 in 2007 dollars (SSA data)	

Table 8

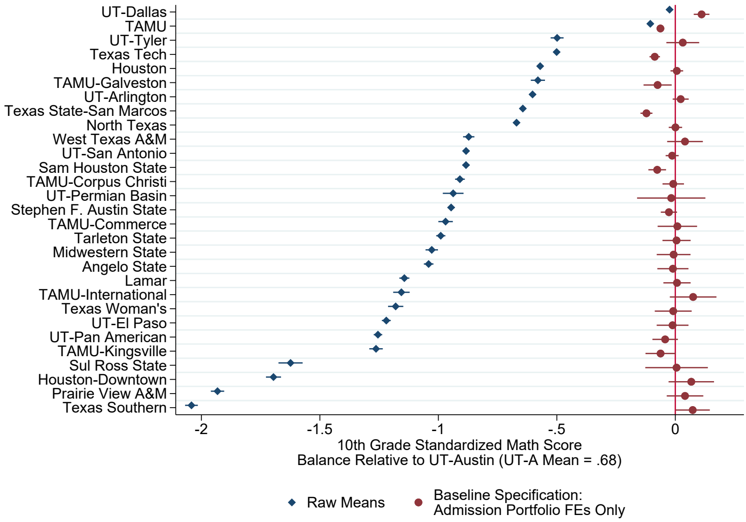
Effect of School Characteristics on 2007 Earnings (Black and Hispanic Students Only, 1989 Cohort)

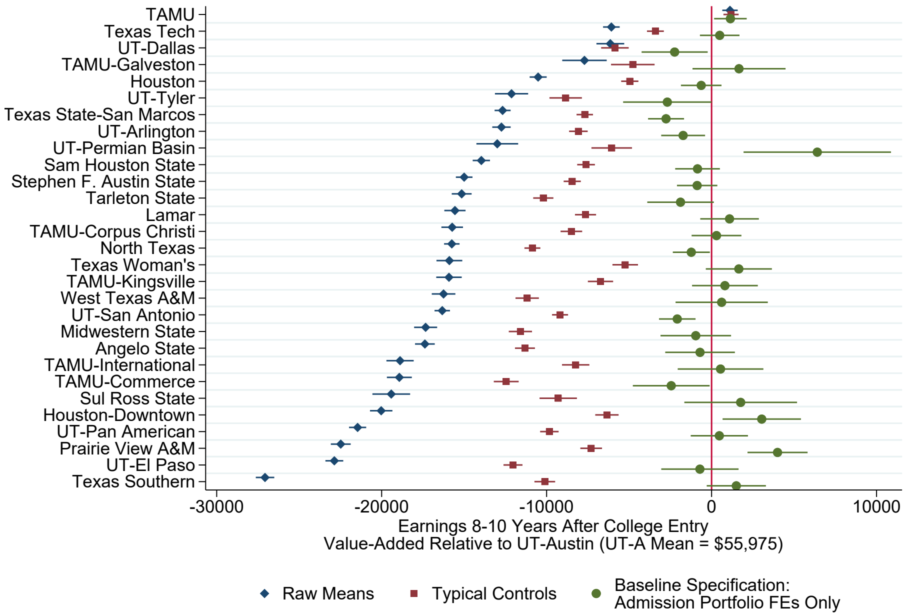
	School SAT score/100		Log net tuition		Barron's index	
Dependent variable	Basic	Self-revelation	Basic	Self-revelation	Basic	Self-revelation
All black and Hispanic students						
Parameter estimate for	0.067	0.076	0.173	0.138	0.063	0.049
effect of quality measure	(.019)	(.032)	(.056)	(.071)	(.022)	(.036)
on log 2007 earnings	{.028}	{.042}	{.076}	{ .092}	{.033}	{.046}
Sample size	1,508		1,508		1,508	
All black and Hispanic students, excluding historically black colleges and universities						
Parameter estimate for	0.122	0.120	0.187	0.116	0.158	0.143
effect of quality measure	(.030)	(.042)	(.064)	(.079)	(.040)	(.053)
on log 2007 earnings	{.035}	{.056}	{.081}	{.101}	{.038}	{.051}
Sample size	995		995		995	

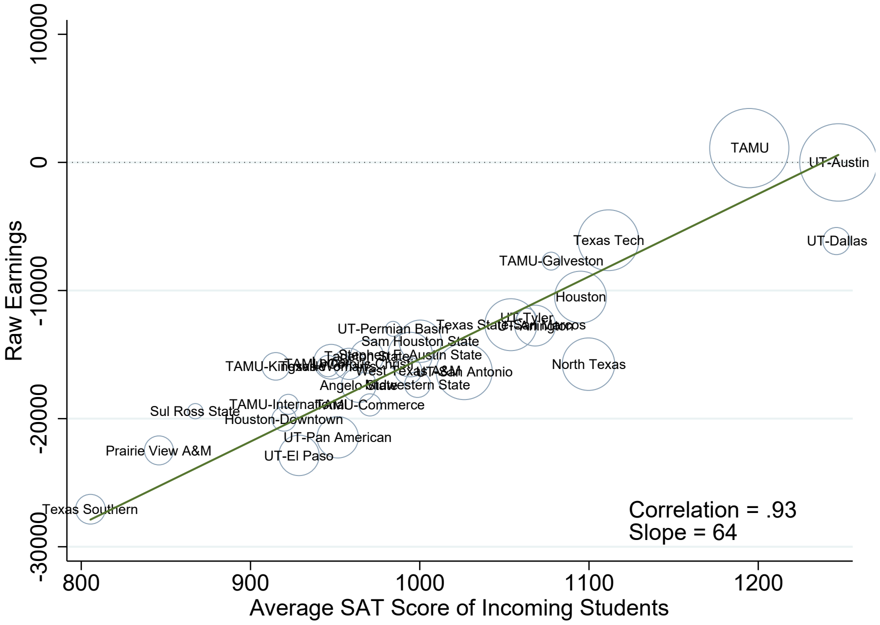
Updating Dale/Krueger: Mountjoy and Hickman (2020)

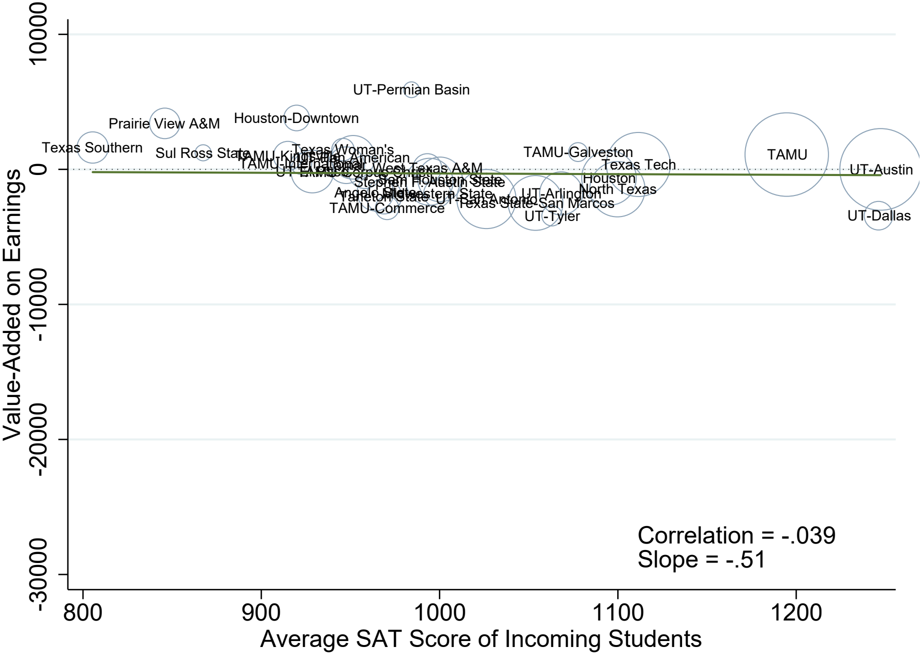
- A recent paper by Mountjoy and Hickman (2020) updates the Dale/Krueger strategy using administrative data from Texas
- Rather than looking at overall return to selectivity, estimate a "value-added" model with a different effect for every college, conditioning on DK application/admission controls
- Relate college value-added to selectivity and other institution characteristics
- Consistent with DK, Mountjoy and Hickman find limited returns to selectivity
- Estimated college value-added is positively correlated with other inputs like instructional expenditures and faculty/student ratio

Figure 3: Validating the Matched Applicant Approach: Ability Balance across College Treatments



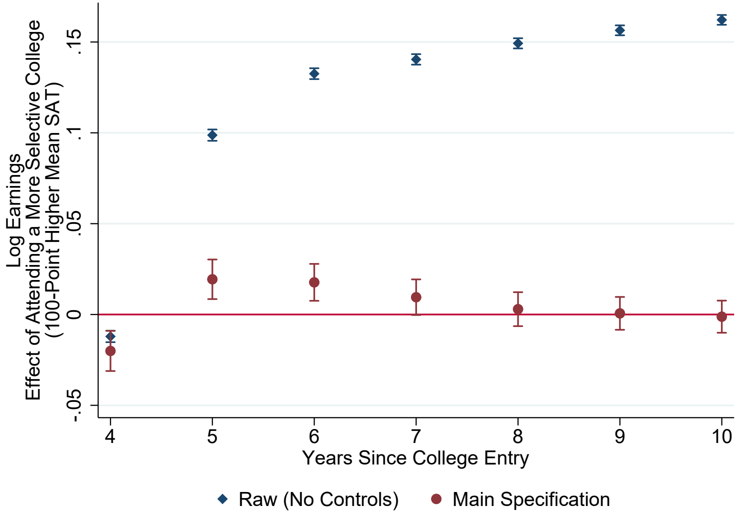

TAMU
Texas Tech
UT-Dallas
TAMU-Galveston
Houston
UT-Tyler
Texas State-San Marcos
UT-Arlington
UT-Permian Basin
Sam Houston State
Stephen F. Austin State
Tarleton State
Lamar
TAMU-Corpus Christi
North Texas
Texas Woman's
TAMU-Kingsville
West Texas A&M
UT-San Antonio
Midwestern State
Angelo State
TAMU-International
TAMU-Commerce
Sul Ross State
Houston-Downtown
UT-Pan American
Prairie View A&M
UT-El Paso
Texas Southern
-30000
-20000
-10000
0
10000
Earnings 8-10 Years After College Entry
Value-Added Relative to UT-Austin (UT-A Mean = $55,975)
Raw Means
Typical Controls
Baseline Specification:
Admission Portfolio FEs Only

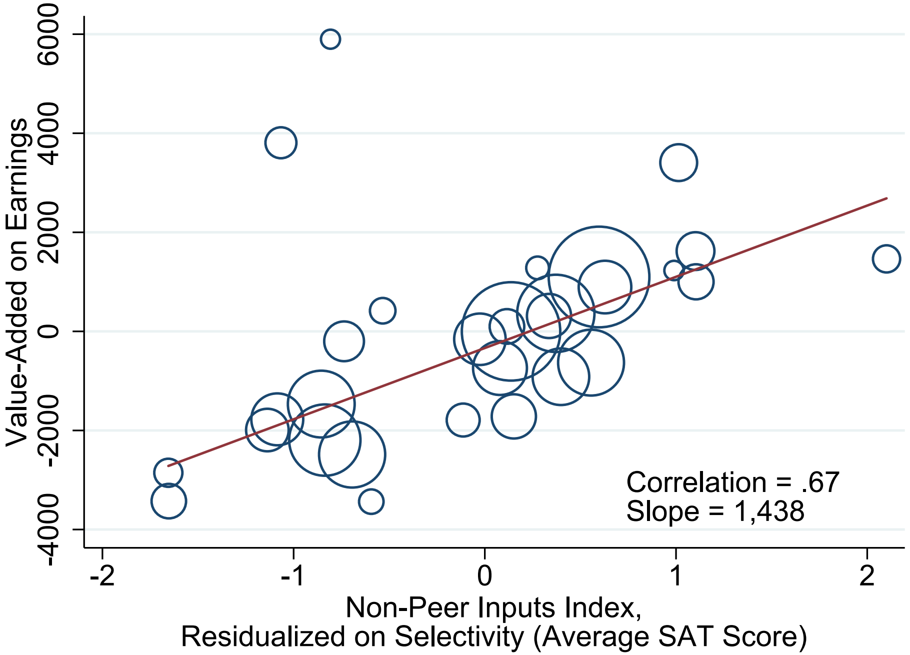

Raw Earnings
Average SAT Score of Incoming Students
10000
0
-10000
-20000
-30000
800
900
1000
1100
1200
TAMU
UT-Austin
Texas Tech
UT-Dallas
TAMU-Galveston
Houston
UT-Tyler
UT-Arlington
Texas State-San Marcos
UT-Permian Basin
Sam Houston State
Stephen F. Austin State
North Texas
UT-San Antonio
Midwestern State
TAMU-Commerce
Sul Ross State
Houston-Downtown
UT-Pan American
Prairie View A&M
UT-El Paso
Texas Southern
Correlation = .93
Slope = 64



Value-Added on Earnings
10000
0
-10000
-20000
-30000
800
900
1000
1100
1200
Average SAT Score of Incoming Students
UT-Permian Basin
Houston-Downtown
Prairie View A&M
Texas Southern
Sul Ross State
Texas Woman's
TAMU-Galveston
TAMU
Texas Tech
UT-Austin
Houston
North Texas
UT-Arlington
Texas State-San Marcos
UT-Tyler
UT-Dallas
TAMU-Commerce
West Texas A&M
Stephen F. Austin State
Correlation = -.039
Slope = -.51

Figure 7: Early Career Dynamics of the Return to College Selectivity





Value-Added on Earnings
6000
4000
2000
0
-2000
-4000
-2
-1
0
1
2
Non-Peer Inputs Index,
Residualized on Selectivity (Average SAT Score)
Correlation = .67
Slope = 1,438

References

- Acemoglu, D., and Autor, D. (2011). "Skills, tasks, and technologies: implications for employment and earnings." *Handbook of Labor Economics*, Volume 4b.
- Angrist, J., Imbens, G., and Rubin, D. (1996). "Identification of causal effects using instrumental variables." *Journal of the American Statistical Association* 91(434).
- Angrist, J., and Krueger, A. (1991). "Does compulsory school attendance affect schooling and earnings?" *Quarterly Journal of Economics* 106(4).
- Angrist, J., and Pischke, J. (2009). Mostly Harmless Econometrics: An Empiricist's Compansion. Princeton University Press.
- Calonico, S., Cattaneo, M., and Titiunik, R. (2014). "Robust nonparametric confidence intervals for regression-discontinuity designs." *Econometrica* 82(6).
- Card, D. (1999). "The causal effect of education on earnings." *Handbook of Labor Economics*, Volume 3.
- Chetty, R., Friedman, J., Saez, E., Turner, N., and Yagan, D. (2020). "Income segregation and intergenerational mobility across colleges in the United States." *Quarterly Journal of Economics* 135(3).
- Clark, D., and Martorell, P. (2014). "The signaling value of a high school diploma." *Journal of Political Economy* 122(2).
- Dale, S., and Krueger, A. (2002). "Estimating the payoff to attending a more selective college: an application of selection on observables and unobservables." *Quarterly Journal of Economics* 117(4).

References

- Dale, S., and Krueger, A. (2014). "Estimating the effects of college characteristics over the career using administrative earnings data." *Journal of Human Resources* 49(2).
- Goldin, C., and Katz, L. (2008). The Race Between Education and Technology. Harvard University Press.
- Holland, P. (1986). "Statistics and causal inference." *Journal of the American Statistical Association* 81(396).
- Imbens, G., and Angrist, J. (1994). "Identification and estimation of local average treatment effects." *Econometrica* 62(2).
- Imbens, G., and Kalyanaraman, K. (2011). "Optimal bandwidth choice for the regression discontinuity estimator." *Review of Economic Studies* 79(3).
- Imben, G., and Lemieux, T. (2008). "Regression discontinuity designs: a guide to practice." Journal of Econometrics 142(2).
- McCrary, J. (2008). "Manipulation of the running variable in the regression discontinuity design: a density test." *Journal of Econometrics* 142(2).
- Mountjoy, J., and Hickman, B. (2020). "The returns to college(s): estimating value-added and match effects in higher education." Working paper.
- Spence, M. (1973). "Job market signaling." *Quarterly Journal of Economics* 87(3).
- Zimmerman, S. (2014). "The returns to college admission for academically marginal students." *Journal of Labor Economics* 32(4).

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture 2 - Self-Selection

Chris Walters

University of California, Berkeley and NBER

Self-selection

- Classic idea in labor economics: Self-selection
- Individuals choose between opportunities based on heterogeneous unobserved returns
- We've already encountered some versions of this in the context of the returns to schooling
- Applications: educational choice, occupational choice, labor force participation, immigration
- We'll start with general discussion of self-selection models and related econometrics, then look at some applications

Roy Model

- Roy (1951) sought to understand the influence of occupational choice on the observed distribution of earnings
- Consider individuals indexed by i choosing a binary variable $D_i \in \{0, 1\}$ indicating occuption, e.g. hunting vs. fishing
- Y_{i1} and Y_{i0} are i's potential earnings associated with each choice
- Realized earnings are $Y_i = Y_{i0} + (Y_{i1} - Y_{i0})D_i$
- Pure Roy (1951) model: Individuals want to maximize Y_i, so choose the occupation with the best potential outcome:

$$D_i = 1\{Y_{i1} > Y_{i0}\}$$

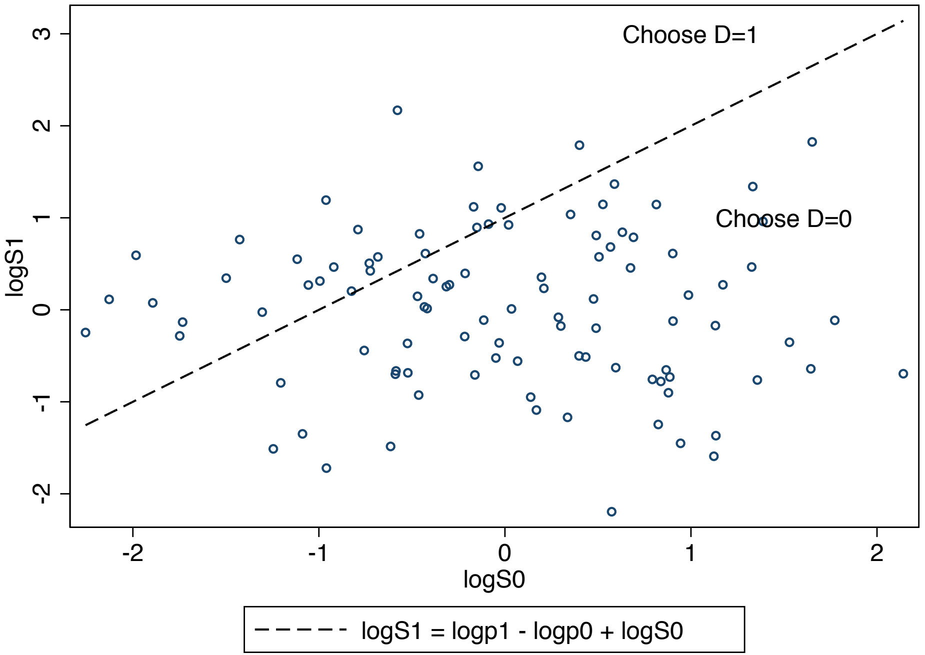
Roy Model

- Questions of interest:
  - Will the best hunters hunt?
  - Will the best fishermen/women fish?
- Suppose potential earnings are given by

$$Y_{id} = p_d S_{id}, \;\; d \in \{0, 1\}$$

- S_{id} is skill in occupation d, and p_d is price of output
- A worker who is indifferent between the two occupations satisfies

$$\log S_{i1} = \log p_0 - \log p_1 + \log S_{i0}$$


Choose D=1
Choose D=0
logS1
logS0
3
2
1
0
-1
-2
-2
-1
0
1
2
logS1 = logp1 - logp0 + logS0

Roy Model: Special Cases

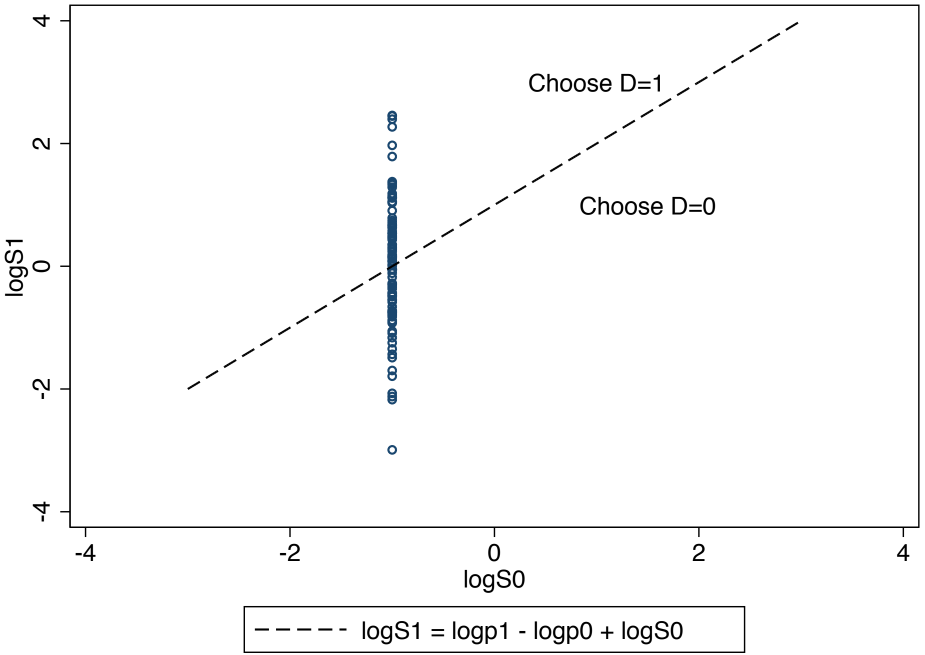
- Suppose there is no variation in potential earnings in sector 0, so $S_{i0} = \bar{S}_0 \ \forall i$

- In this case the decision rule is

$$D_i = 1\left\{S_{i1} \geq \left(\frac{p_0}{p_1}\right)\bar{S}_0\right\}$$

- Those with the most skill in sector 1 choose $D_i = 1$

- Everyone with $D_i = 1$ earns more than anyone with $D_i = 0$



Choose D=1
Choose D=0
logS1
logS0
-4
-2
0
2
4
logS1 = logp1 - logp0 + logS0

Roy Model: Special Cases

- Suppose we have perfect correlation between $\log S_{i0}$ and $\log S_{i1}$:

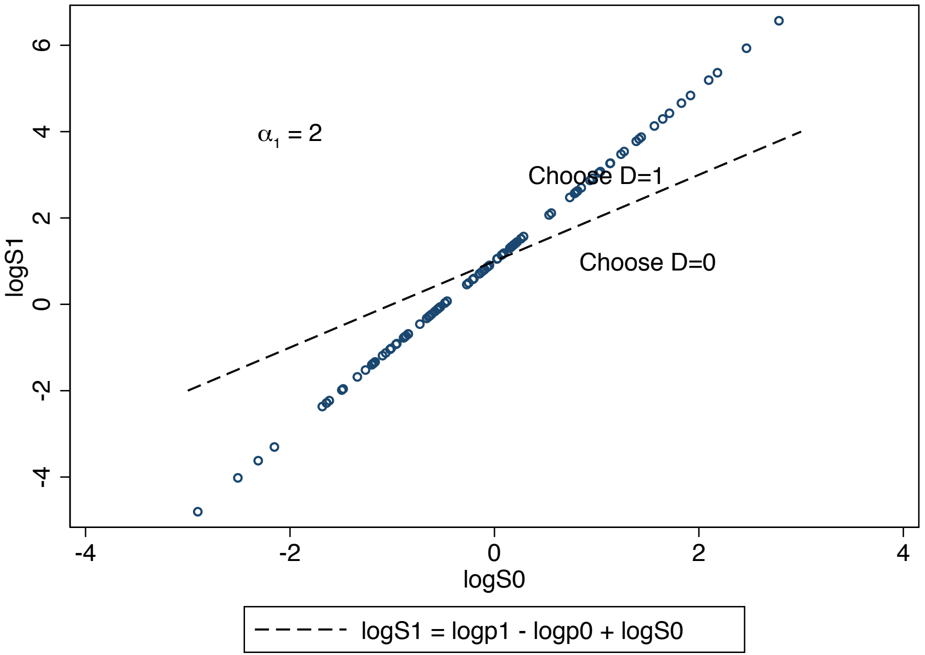
$$\log S_{i1} = \alpha_0 + \alpha_1 \log S_{i0}, \ \ \alpha_1 > 0$$

- This is a one-factor model
- Decision rule:

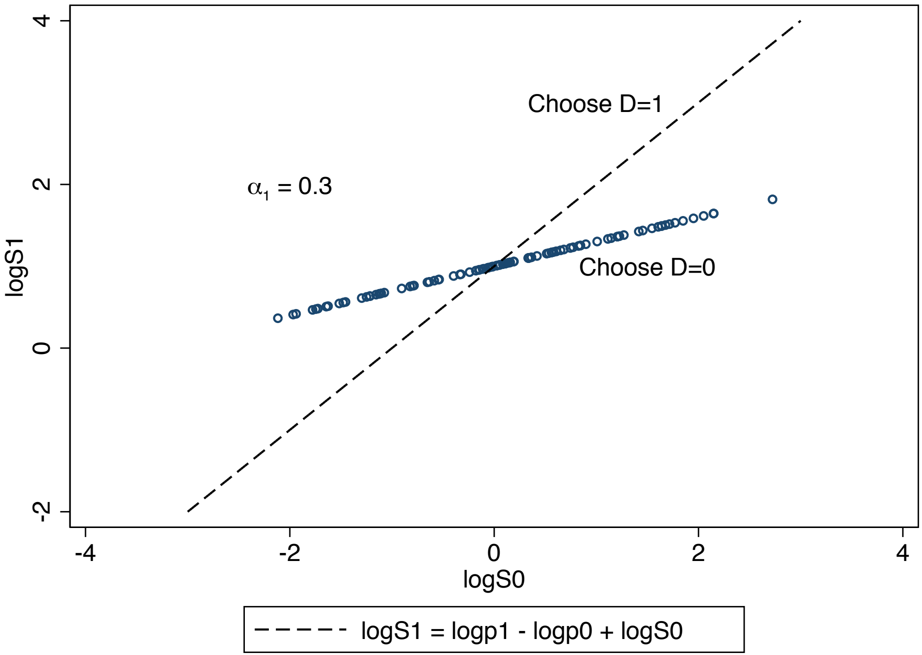
$$D_i = 1\{\alpha_0 + \alpha_1 \log S_{i0} \geq \log p_0 - \log p_1 + \log S_{i0}\}$$

$$= 1\{(\alpha_1 - 1)\log S_{i0} \geq \log p_0 - \log p_1 - \alpha_0\}$$

- Higher skilled choose $D_i = 1$ iff $\alpha_1 \geq 1$
- Note that $Var(\log S_{i1}) = \alpha_1^2 Var(\log S_{i0})$. Higher skilled choose the sector with higher variance of skill



$\alpha_1 = 2$
Choose D=1
Choose D=0
logS1
logS0
6
4
2
0
-2
-4
-4
-2
0
2
4
logS1 = logp1 - logp0 + logS0



Choose D=1
α1 = 0.3
Choose D=0
logS1
logS0
4
2
0
-2
-4
-2
0
2
4
logS1 = logp1 - logp0 + logS0

Generalized Roy Model

- Generalized Roy model (Eisenhauer et al., 2015): Preference for alternative d depends on Y_{id} as well as a heterogeneous cost C_{id}:

$$D_i = 1\left\{Y_{i1} - C_{i1} > Y_{i0} - C_{i0}\right\}$$

- Allows us to ask richer questions about selection on both levels and gains
  - Is average Y_{i1} higher or lower for individuals that choose $D_i = 1$?
  - Is average Y_{i0} higher or lower for individuals that choose $D_i = 1$?
  - Are average *gains* $Y_{i1} - Y_{i0}$ larger or smaller for individuals that choose $D_i = 1$?
- Close link between generalized Roy model and econometric models of treatment effect heterogeneity

Self-Selection Example: Labor Supply

- Simple example of a selection model: Labor supply problem

$$\max_{c,h} c - v(h) \text{ s.t. } c \leq wh + V$$

- At interior solutions:

$$v'(h^*) = w$$

- At corner solutions:

$$v'(0) > w$$

- Reservation wage is $w^* = v'(0)$; work if $w \geq w^*$

Labor Supply Selection

- Suppose individuals' reservation wages are described by

$$w_i^* = X_i'\theta + \eta_i$$

- Offered wages are

$$w_i = X_i'\beta + \epsilon_i$$

- Assume $E[\eta_i|X_i] = E[\epsilon_i|X_i] = 0$, so $X_i'\theta$ and $X_i'\beta$ are population CEFs
- Individual i works $(D_i = 1)$ when

$$X_i'\beta + \epsilon_i \geq X_i'\theta + \eta_i$$

$$\iff X_i'(\beta - \theta) + (\epsilon_i - \eta_i) \geq 0$$

$$\iff X_i'\psi \geq v_i$$

Labor Supply Selection

$$D_i = 1\{X_i'\psi \geq v_i\}$$

- $D_i^* = X_i'\psi - v_i$ is a **latent index** determining D_i
- We observe outcomes in the sample with $D_i = 1$. CEF in this sample is

$$E[w_i|X_i, D_i = 1] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

- If ϵ_i and v_i are independent, the last term is $E[\epsilon_i|X_i] = 0$ and OLS recovers β
- This is equivalent to saying we have a random sample – selection into the sample is unrelated to outcomes
- If ϵ_i and v_i aren't independent, we'll have $E[\epsilon_i|X_i, D_i = 1] \neq 0$, and OLS on observed sample is inconsistent

Selection with Normality

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

- Suppose that ϵ_i and v_i are joint normal:

$$(\epsilon_i, v_i)|X_i \sim N\left((0,0), \left[\begin{array}{cc} \sigma_\epsilon^2 & \rho\sigma_\epsilon \\ \rho\sigma_\epsilon & 1 \end{array}\right]\right)$$

- Then we can work out the expected error conditional on $D_i = 1$
- Under normality, conditional expectations are linear:

$$E\left[\epsilon_i|X_i, v_i\right] = \rho\sigma_\epsilon v_i.$$

Mills Ratios

- The CEF of w_i in the observed sample is

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + E\left[\epsilon_i|X_i, v_i < X_i'\psi\right]$$

$$= X_i'\beta + \rho\sigma_\epsilon E\left[v_i|X_i, v_i < X_i'\psi\right]$$

$$= X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\psi\right)$$

- Here $\lambda(x)$ is the conditional expectation of a standard normal random variable truncated from above, also known as the **inverse Mills ratio:**

$$\lambda(x) = -\frac{\phi(x)}{\Phi(x)}.$$

Heckit

$$E\left[w_i|X_i, D_i = 1\right] = X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\psi\right)$$

- ψ can be estimated via a first-step probit of D_i on X_i
- Then run a second-step regression in the $D_i = 1$ sample:

$$w_i = X_i'\beta + \rho\sigma_\epsilon \cdot \lambda\left(X_i'\hat{\psi}\right) + u_i$$

- The Mills ratio is a **control function** or **selection correction** that accounts for selection into the observed sample
- This is Heckman's (1974, 1976, 1979) two-step selection correction ("Heckit")

Heckit Identification

- Suppose X_i is just a constant. Then the second-step regression is

$$w_i = \beta + \rho\sigma_\epsilon \cdot \lambda\left(\hat{\psi}\right) + u_i$$

$$= \delta + u_i$$

- The constant here is $\delta = (\beta + \rho\sigma_\epsilon\lambda(\psi))$, so β and $\rho\sigma_\epsilon$ are not separately identified
- More generally, if outcome and selection equations are saturated in X_i, main effects and Mills ratio term are not separately identified
- This is unattractive – there is typically no reason to believe $E\left[w_i|X_i\right]$ is linear in X_i

Heckit Identification

- Solution: Suppose there are additional variables Z_i in the selection equation, so

$$D_i = 1\left\{X_i'\psi + Z_i'\pi > v_i\right\}$$

- Assume $E\left[\epsilon_i | X_i, Z_i\right] = 0$. Then second-step CEF is

$$E\left[w_i | X_i, Z_i, D_i = 1\right] = X_i'\beta + \rho\sigma_\epsilon\lambda\left(X_i'\psi + Z_i'\pi\right)$$

- If $\pi \neq 0$ this can be estimated even if X_i is saturated since variation in Z_i separately identifies the selection term
- Identifying a Heckit without relying on functional form restrictions requires finding a Z_i that shifts the probability of selection but is excludable from the outcome equation
- Sound familiar?

Heckit with Instruments

- The requirements for a good Z_i in the Heckit model are the same as the requirements for a good instrument when we're doing IV
- This is not a coincidence. Control function and IV are methods for solving the same problem

Selection and Treatment Effects

- To see the connection between control function and IV, consider a heterogeneous treatment effects model:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- Here $\alpha_d = E[Y_i(d)]$ so $E[\epsilon_{id}] = 0$
- If we had random samples of $Y_i(1)$ and $Y_i(0)$ we could run OLS (i.e., take means) and estimate $ATE = \alpha_1 - \alpha_0$

Selection and Treatment Effects

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- But we only observe $Y_i(1)$ when $D_i = 1$, and we only observe $Y_i(0)$ when $D_i = 0$
- These are not random samples if treatment is not as good as randomly assigned
- We therefore have sample selection problems for both $Y_i(1)$ and $Y_i(0)$
- Treatment effects estimation is a two-sided sample selection problem
- An instrument is needed to solve this problem

IV and Selection Models

- We have seen that IV and control function are two methods for solving the same problem
- How should we think about the relationship between parametric sample selection models and the nonparametric LATE model of Imbens and Angrist (1994)?
- How should we think about the relationship between estimates produced by IV and control function?

IV and Selection Models

- To better understand the relationships between latent index models and the LATE model, consider a treatment effects model with a binary treatment and binary instrument:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

- Suppose selection into the $D_i = 1$ sample follows the rule

$$D_i = 1\{\psi_0 + \psi_1 Z_i > v_i\}$$

$$(\epsilon_{i1}, \epsilon_{i0}, v_i) \perp\!\!\!\perp Z_i$$

$$v_i \sim F(v)$$

- $F(v)$ is some strictly increasing parametric distribution function (e.g. the normal CDF)

IV and Selection Models

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

$$D_i = 1\{\psi_0 + \psi_1 Z_i > v_i\}$$

$$(\epsilon_{i1}, \epsilon_{i0}, v_i) \perp\!\!\!\perp Z_i$$

$$v_i \sim F(v)$$

- This selection model appears to be more restrictive than the LATE model, which involves no distributional assumptions

LATE Model and Selection Model: Equivalence

- Vytlacil (2002) shows that this selection model is the LATE model, in the sense that
  - The selection model satisfies the LATE assumptions
  - The LATE assumptions imply that the selection model rationalizes the observed and counterfactual outcomes and treatments

LATE Model and Selection Model: Equivalence

- The first part of the proof is straightforward. Note that

$$Y_i(0) = \alpha_0 + \epsilon_{i0},\ Y_i(1) = \alpha_1 + \epsilon_{i1},$$

$$D_i(0) = 1\{\psi_0 > v_i\}\ ,\ D_i(1) = 1\{\psi_0 + \psi_1 > v_i\}$$

- $Y_i(d)$ and $D_i(z)$ are functions of $(\epsilon_{i0}, \epsilon_{i1}, v_i)$ which are independent of Z_i, so independence/exclusion are satisfied

- If $\psi_1 > 0$, then $D_i(1) \geq D_i(0)$ and monotonicity is satisfied

- $Pr[D_i(1) > D_i(0)] = Pr[\psi_0 + \psi_1 > v_i \geq \psi_0] > 0$ since $F(\cdot)$ is strictly increasing, so there is a first stage

- The selection model therefore satisfies the assumptions of the LATE framework

LATE Model and Selection Model: Equivalence

- To show that the LATE model implies the selection model representation, first note that with a binary Z_i the "parametric" assumption $v_i \sim F(v)$ is not really a restriction
- For any strictly increasing distribution function $G(\cdot)$ we can write

$$D_i = 1\left\{G^{-1}\left(F\left(\psi_0 + \psi_1 Z_i\right)\right) > G^{-1}(F(v_i))\right\}$$

$$= 1\left\{\tilde{\psi}_0 + \tilde{\psi}_1 Z_i > \tilde{v}_i\right\},$$

- where

$$\tilde{\psi}_0 = G^{-1}\left(F(\psi_0)\right),\ \tilde{\gamma}_1 = G^{-1}(F(\psi_0 + \psi_1)) - G^{-1}(F(\psi_0))$$

$$\tilde{v}_i = G^{-1}(F(v_i))$$

LATE Model and Selection Model: Equivalence

$$D_i = 1\left\{\tilde{\psi}_0 + \tilde{\psi}_1 Z_i > \tilde{v}_i\right\},$$

- The new selection error $\tilde{v}_i = G^{-1}(F(v_i))$ has CDF $G(\cdot)$
- The same selection model can be represented with any distribution function
- It is therefore sufficient to show that the LATE model implies a selection model representation for SOME distribution function

LATE Model and Selection Model: Equivalence

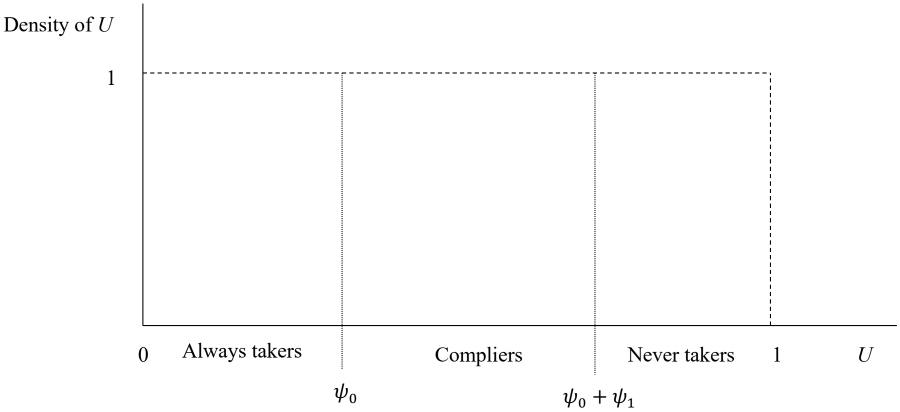
- Let $u_i \sim U(0,1)$ be independent of Z_i, and define

$$U_i = \begin{cases} u_i \times Pr\left[D_i(0)=1\right], & D_i(0)=1 \\ Pr\left[D_i(0)=1\right] + u_i \times Pr\left[D_i(1)>D_i(0)\right], & D_i(1)>D_i(0) \\ Pr\left[D_i(1)=1\right] + u_i \times Pr\left[D_i(1)=0\right], & D_i(1)=0 \end{cases}$$

- Then we can write

$$D_i = 1\left\{\psi_0 + \psi_1 Z_i > U_i\right\}$$

- Here $\psi_0 = Pr\left[D_i(0)=1\right]$, $\psi_1 = Pr\left[D_i(1)>D_i(0)\right]$, and $U_i \sim U(0,1)$



Density of U
1
0
Always takers
Compliers
Never takers
1
U
ψ_0
$\psi_0 + \psi_1$

LATE Model and Selection Model: Equivalence

- U_i is uniform on $(0, \psi_0)$ for always takers, on $(\psi_0, \psi_0 + \psi_1)$ for compliers, and on $(\psi_0 + \psi_1, 1)$ for never takers
- This model implies the same observed and counterfactual treatment choices and outcomes as the LATE model
- We can equivalently represent the selection model with the distribution $F(\cdot)$ by applying $F^{-1}(\cdot)$ to both sides of the treatment selection equation
- We have therefore shown that the LATE model and the selection model are equivalent: They are two ways of representing the same information
- Vytlacil (2002) shows that this applies to the more general LATE model with multiple instruments

IV and Control Function

- Selection model with uniform representation of selection error:

$$Y_i(1) = \alpha_1 + \epsilon_{i1}$$

$$Y_i(0) = \alpha_0 + \epsilon_{i0}$$

$$D_i = 1\{\psi_0 + \psi_1 Z_i > U_i\}$$

$$U_i \sim U(0,1)$$

$$(\epsilon_{i1}, \epsilon_{i0}, U_i) \perp\!\!\!\perp Z_i$$

- We've shown that this is the LATE model
- Does this mean that IV and control function estimates of treatment effects are also equivalent?

IV and Control Function

- No. In fact, we cannot estimate this model by control function without further assumptions
- To form control functions we need to specify $E[\epsilon_{id}|U_i]$, which we haven't done
- Control function yields estimates of α_1 and α_0, and therefore the *ATE* $\alpha_1 - \alpha_0$
- The *ATE* is not identified in the LATE model – we can only get the *LATE*
- We have to assume more if we want to extrapolate from *LATE* to *ATE*

Complier Potential Outcomes

- To understand control function extrapolation, it's useful to start with what is nonparametrically identified in the LATE framework
- We know the average treatment effect for compliers is identified (LATE theorem)
- It turns out that other features of complier potential outcomes are identified as well (Imbens and Rubin, 1997; Abadie, 2003)
- Individuals with $D_i = Z_i = 1$ are a mix of always takers and compliers:

$$E[Y_i|D_i = Z_i = 1] = \left(\frac{\pi_{AT}}{\pi_{AT} + \pi_C}\right) E\left[Y_i(1)|AT\right] + \left(\frac{\pi_C}{\pi_{AT} + \pi_C}\right) E\left[Y_i(1)|C\right]$$

Complier Potential Outcomes

- Always taker outcome is observed directly as

$$E[Y_i|D_i = 1, Z_i = 0] = E[Y_i(1)|AT]$$

- Population shares are also identified since

$$\pi_{AT} = \Pr[D_i = 1|Z_i = 0]$$

$$\pi_C = \Pr[D_i = 1|Z_i = 1] - \Pr[D_i = 1|Z_i = 0]$$

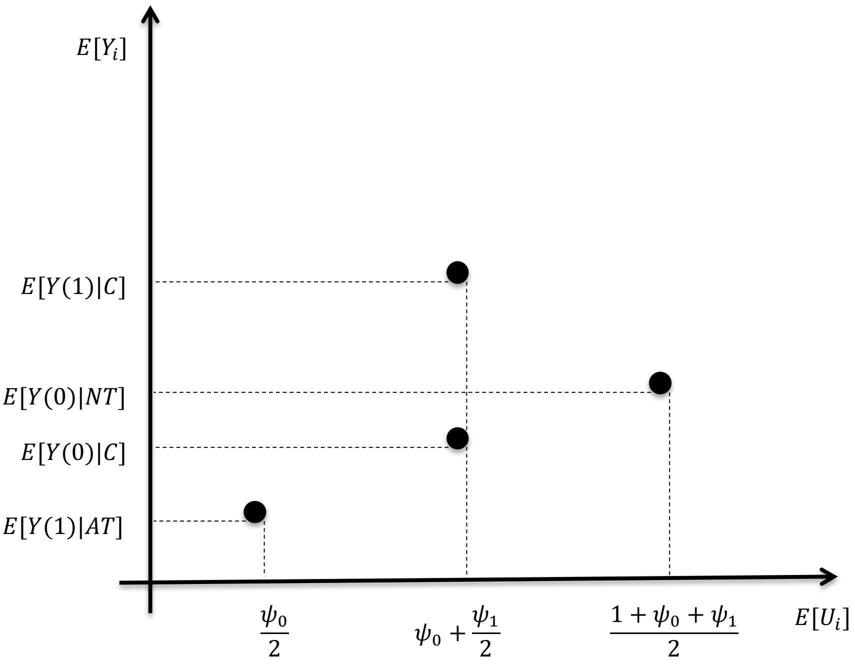
- We can then back out the average complier $Y_i(1)$ as

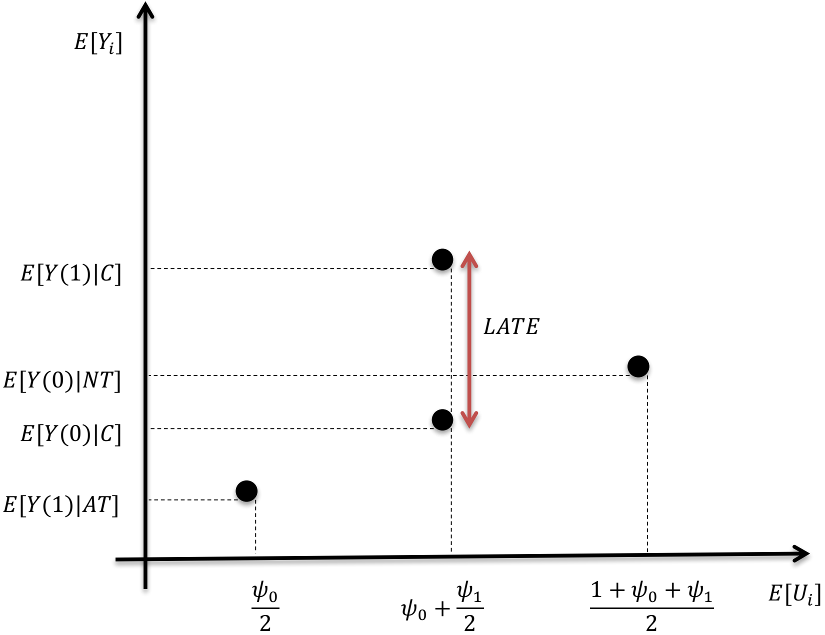
$$E[Y_i(1)|C] = \left(\frac{\pi_{AT} + \pi_C}{\pi_C}\right) E[Y_i|D_i = Z_i = 1] - \left(\frac{\pi_{AT}}{\pi_C}\right) E\left[Y_i(1)|AT\right]$$

- By the same reasoning, we can back out $E[Y_i(0)|C]$ from the complier/never taker mix with $D_i = Z_i = 0$

Control Function Extrapolation

- In the LATE framework we can identify:
  - $E[Y_i(1)|AT]$
  - $E[Y_i(0)|NT]$
  - $E[Y_i(1)|C]$
  - $E[Y_i(0)|C]$
- We can therefore identify means of $Y_i(1)$ and $Y_i(0)$ for two groups each
- In selection model notation, this yields two points on the curve $E[Y_i(d)|U_i]$ for each potential outcome

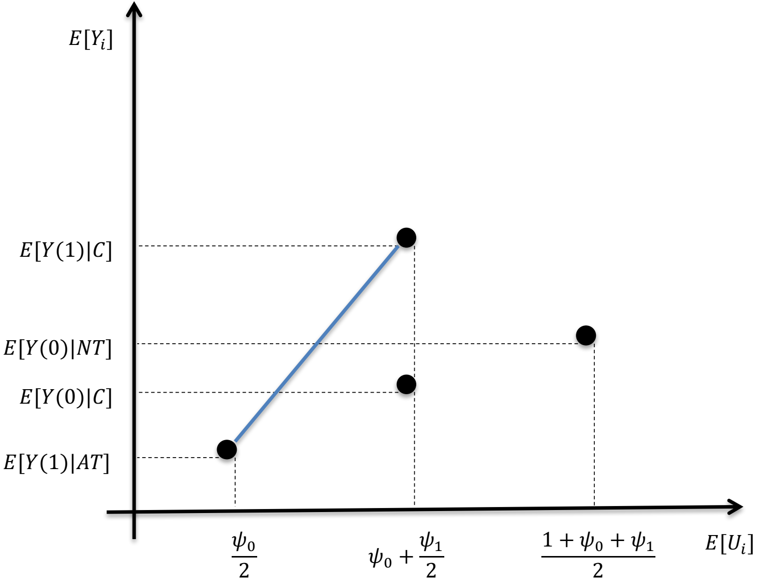

$E[Y_i]$
$E[Y(1)|C]$
$E[Y(0)|NT]$
$E[Y(0)|C]$
$E[Y(1)|AT]$
$\frac{\psi_0}{2}$
$\psi_0 + \frac{\psi_1}{2}$
$\frac{1 + \psi_0 + \psi_1}{2}$
$E[U_i]$



$E[Y_i]$
$E[Y(1)|C]$
$LATE$
$E[Y(0)|NT]$
$E[Y(0)|C]$
$E[Y(1)|AT]$
$\frac{\psi_0}{2}$
$\psi_0 + \frac{\psi_1}{2}$
$\frac{1 + \psi_0 + \psi_1}{2}$
$E[U_i]$

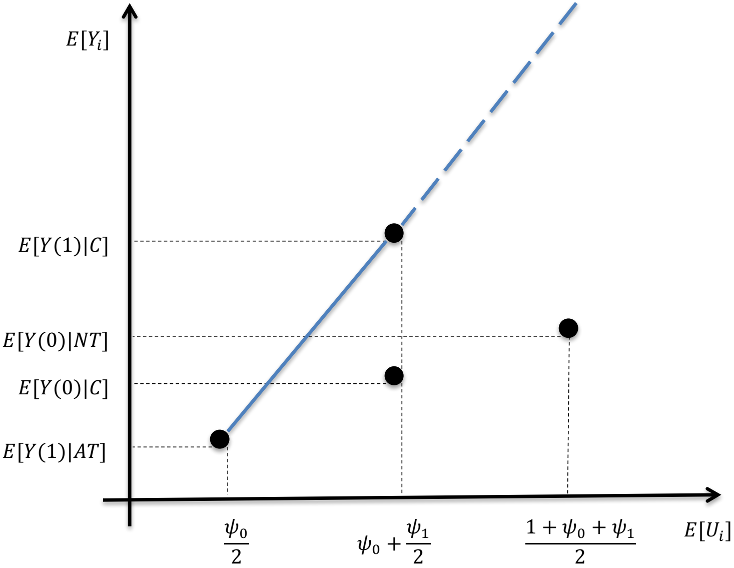
Extrapolation from LATE

- Without further assumptions we cannot identify any other treatment effects
- But by specifying a functional form for $E\left[Y_i(d)|U_i\right]$, we can "connect the dots" and extrapolate to predict effects for always takers and never takers
- This allows us to predict the effects of policies that affect different subpopulations than the instrument at hand
- More generally, think of selection model as a device for extrapolating from available research design to predict impacts of other experiments



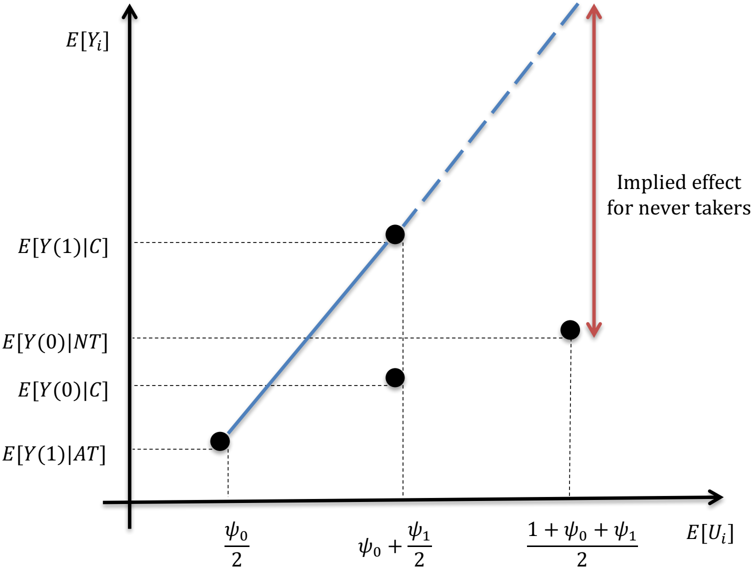
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



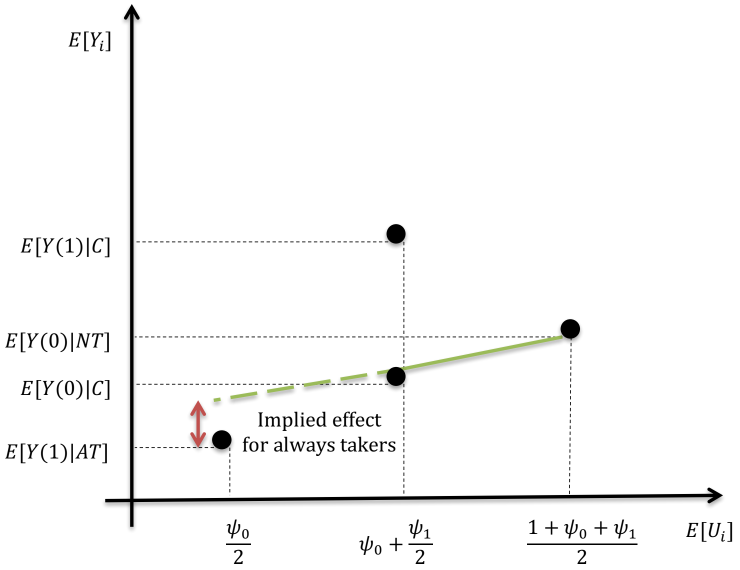
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



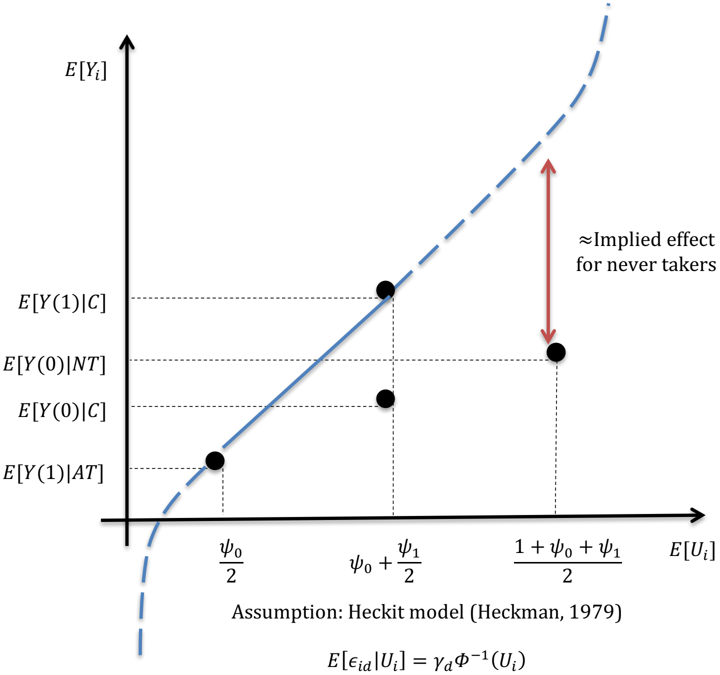
Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



Assumption: Linear selection (Olsen, 1980)

$$E[\epsilon_{id}|U_i] = \gamma_d U_i$$



$E[Y_i]$
$E[Y(1)|C]$
$E[Y(0)|NT]$
$E[Y(0)|C]$
$E[Y(1)|AT]$
≈Implied effect for never takers
$\frac{\psi_0}{2}$
$\psi_0 + \frac{\psi_1}{2}$
$\frac{1 + \psi_0 + \psi_1}{2}$
$E[U_i]$
Assumption: Heckit model (Heckman, 1979)
$E[\epsilon_{id}|U_i] = \gamma_d \Phi^{-1}(U_i)$

Marginal Treatment Effects

- Letting $U_i \sim U(0,1)$, choosing $E\left[Y_i(d)|U_i\right]$ implies a functional form for **marginal treatment effects** (MTE):

$$MTE(u) = E\left[Y_i(1) - Y_i(0)|U_i = u\right]$$

- MTEs are average treatment effects for individuals at a particular percentile of the unobserved cost of taking treatment (Heckman et al., 1999, 2005, 2006; Carneiro et al., 2009, 2010)

- $MTE(u)$ can be thought of as the $LATE$ associated with a hypothetical instrument that shifts the probability of treatment from u to $u + \Delta$ for small Δ

- With a continuous instrument, MTEs can be estimated as derivatives of average Y_i with respect to the conditional probability of treatment (local IV; Heckman and Vytlacil, 1999)

- With a discrete instrument, estimation involves extrapolation/interpolation from available LATEs (Brinch et al., 2017)

Marginal Treatment Effects

- Many treatment effects of interest can be defined as weighted averages of MTEs – useful for thinking about external validity:

$$\int_0^1 \omega(u) MTE(u) du$$

- Let $\pi(z) = Pr\left[D_i = 1 | Z_i = z\right]$, and $p = Pr\left[Z_i = 1\right]$
- Weights for notable treatment effects:

$$ATE: \ \omega(u) = 1$$

$$TOT: \ \omega(u) = \frac{p1\left\{u < \pi(1)\right\} + (1-p)1\left\{u < \pi(0)\right\}}{\pi(1)p + \pi(0)(1-p)}$$

$$TNT: \ \omega(u) = \frac{p1\left\{u \geq \pi(1)\right\} + (1-p)1\left\{u \geq \pi(0)\right\}}{(1-\pi(1))p + (1-\pi(0))(1-p)}$$

$$LATE: \ \omega(u) = \frac{1\left\{\pi(0) \leq u < \pi(1)\right\}}{\pi(1) - \pi(0)}$$

MTE and Policy Counterfactuals

- Models for MTE can be used to predict the effects of policies that have not been implemented
- Example: Suppose an experiment reduces the price of purchasing health insurance from p_0 to p_1, and the probability of purchase rises from π_0 to π_1
- Individuals with $U_i = \pi_1$ are on the margin between purchasing and not purchasing – we might expect them to purchase in response to a further price cut
- Heckit prediction of effect for marginal population:

$$\widehat{MTE(\pi_1)} = \hat{\alpha}_1 - \hat{\alpha}_0 + (\hat{\gamma}_1 - \hat{\gamma}_0)\,\Phi^{-1}(\hat{\pi}_1)$$

- Can also use estimates of MTEs to predict TOT, TNT, ATE, or effects of other hypothetical policies

Through the Looking Glass

- CF estimate of LATE:

$$\widehat{LATE} = \hat{\alpha}_1 - \hat{\alpha}_0 + \hat{E}\left[\epsilon_{i1} - \epsilon_{i0} | \psi_0 \leq U_i < \psi_0 + \psi_1\right]$$

- In the binary treatment/binary instrument case with two-sided non-compliance, the two-step estimate of LATE produced by any parametric selection model is algebraically equal to the IV estimate (Kline and Walters, 2019)

- The CF estimator exactly fits the IV estimates of mean potential outcomes regardless of functional form – it connects the dots in sample

- In binary/binary case IV and CF coincide when both are used to estimate LATE

  - Equivalence serves as a natural benchmark for assessing overidentified selection models

- The assumption for $E\left[\epsilon_{id} | U_i\right]$ only matters when it is used to predict treatment effects for other subpopulations

When to Extrapolate?

- When is it reasonable to extrapolate from LATE and predict the effects of new policies?
- It depends on the interpretation of U_i, and hence on the instrument
- Equivalent to asking: when is the relationship between always taker/complier $Y_i(1)$'s likely to be a reliable guide to the relationship between complier/never taker $Y_i(1)$'s?
- If Z_i is a price shift, U_i may be viewed as (minus) willingness to pay and extrapolation may be sensible
- What would extrapolation mean in other IV examples?

Selection into Preschool: Kline and Walters (2016)

- Selection model example: Kline and Walters (QJE 2016) investigate effect heterogeneity with respect to counterfactual treatment choices
- Setting: Randomized evaluation of Head Start program
  - Public preschool for disadvantaged children
  - Largest preschool program in the US
  - Basic experimental impacts less impressive than earlier non-experimental analyses of HS
  - But alternative publicly subsidized preschools are now widely available for HS-eligible children. Are effects larger for kids who would otherwise stay home?

TABLE II

EXPERIMENTAL IMPACTS ON TEST SCORES

	Three-year-old cohort			Four-year-old cohort			Cohorts pooled		
Time period	(1) Reduced form	(2) First stage	(3) IV	(4) Reduced form	(5) First stage	(6) IV	(7) Reduced form	(8) First stage	(9) IV
Year 1	0.194	0.699	0.278	0.141	0.663	0.213	0.168	0.682	0.247
	(0.029)	(0.025)	(0.041)	(0.029)	(0.022)	(0.044)	(0.021)	(0.018)	(0.031)
N		1,970			1,601			3,571	

TABLE III

PRESCHOOL CHOICES BY YEAR, COHORT, AND OFFER STATUS

		Offered			Not offered			
Time period	Cohort	(1) Head Start	(2) Other centers	(3) No preschool	(4) Head Start	(5) Other centers	(6) No preschool	(7) C-complier share
Year 1	3-year-olds	0.851	0.058	0.092	0.147	0.256	0.597	0.282
	4-year-olds	0.787	0.114	0.099	0.122	0.386	0.492	0.410
	Pooled	0.822	0.083	0.095	0.136	0.315	0.550	0.338

Kline and Walters (2016): Notation

- $Z_i \in \{0, 1\}$: Randomized experimental offer
- $D_i(z)$: Potential preschool choice.
  - h: Head Start
  - c: Other preschool center
  - n: No preschool
- Monotonicity restriction:

$$D_i(1) \neq D_i(0) \implies D_i(1) = h$$

- People only respond to a Head Start offer by enrolling in Head Start

Kline and Walters (2016): Compliance Groups

- Monotonicity implies that the population can be partitioned into five groups:
  - n-compliers: $D_i(1) = h$, $D_i(0) = n$
  - c-compliers: $D_i(1) = h$, $D_i(0) = c$
  - n-never takers: $D_i(1) = D_i(0) = n$
  - c-never takers: $D_i(1) = D_i(0) = c$
  - Always takers: $D_i(1) = D_i(0) = h$

Kline and Walters (2016): LATE

- The Head Start experiment identifies a LATE:

$$\frac{E\left[Y_i|Z_i=1\right]-E\left[Y_i|Z_i=0\right]}{E\left[1\left\{D_i=h\right\}|Z_i=1\right]-E\left[1\left\{D_i=h\right\}|Z_i=0\right]}$$

$$=E\left[Y_i(h)-Y_i(D_i(0))|D_i(1)\neq D_i(0)\right]$$

$$\equiv LATE_h$$

- This is an effect relative to a mix of counterfactuals:

$$LATE_h=S_c LATE_{ch}+(1-S_c)LATE_{nh}$$

- $LATE_{nh}$ and $LATE_{ch}$ are effects for n and c compliers relative to specific counterfactuals
- S_c is the share of c-compliers among all compliers

Kline and Walters (2016): Policy Relevant Parameters

- $LATE_h$ is the policy-relevant parameter for a marginal expansion of Head Start
  - Policymaker does not control substitution from other programs
  - Not feasible to target policies based on unobserved behavioral responses
  - Effect heterogeneity is not always policy-relevant
- Need a clear motivation for decomposing into "subLATEs" $LATE_{ch}$ and $LATE_{nh}$
  - Scientific interest in understanding small experimental impacts
  - Relevant for policies that change the counterfactual or nature of selection

Kline and Walters (2016): Selection Model

- SubLATEs aren't nonparametrically identified by the experiment
- Estimate via 3-alternative selection model:

$$U_i(h) = \psi_h(X_i, Z_i) + v_{ih}$$

$$U_i(c) = \psi_c(X_i) + v_{ic}$$

$$U_i(n) = 0$$

$$(v_{ih}, v_{ic})|X_i, Z_i \sim N\left(0, \begin{bmatrix} 1 & \rho(X_i) \\ \rho(X_i) & 1 \end{bmatrix}\right)$$

- X_i is a vector of covariates, including demographics and experimental sites

Kline and Walters (2016): Control Functions

- Restrictions on potential outcome CEFs:

$$E\left[Y_i(d)|X_i, Z_i, v_{ih}, v_{ic}\right] = \mu_d(X_i) + \gamma_{dh} v_{ih} + \gamma_{dc} v_{ic}$$

- Averaging over individuals in a particular care alternative gives

$$E\left[Y_i(d)|X_i, Z_i, D_i = d\right] = \mu_d(X_i) + \gamma_{dh}\lambda_h\left(X_i, Z_i, d\right) + \gamma_{dc}\lambda_c\left(X_i, Z_i, d\right)$$

- $\lambda_d(X_i, Z_i, D_i)$ are bivariate versions of the Heckit Mills ratio
- Additive separability between observables and unobservables is key
- Estimates of $\mu_d(x)$, γ_{dh}, and γ_{dc} are used to construct model-based estimates of subLATEs

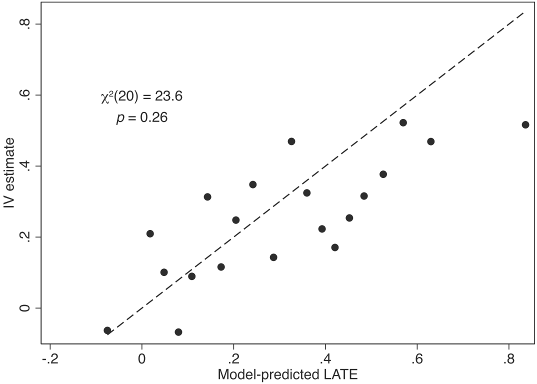


FIGURE 3.—Model-based and IV estimates of LATE. Notes: This figure reproduces Figure A.III from Kline and Walters (2016). The figure is constructed by splitting the Head Start Impact Study sample into vingtiles of the predicted LATE based on the control function estimates reported in Section VIII of the paper. The horizontal axis displays the average predicted LATE in each group, and the vertical axis shows corresponding IV estimates. The dashed line is the 45-degree line. The chi-squared statistic and p-value come from a bootstrap Wald test of the hypothesis that the 45 degree line fits all points up to sampling error. See Appendix F of Kline and Walters (2016) for more details.

TABLE VIII

TREATMENT EFFECTS FOR SUBPOPULATIONS

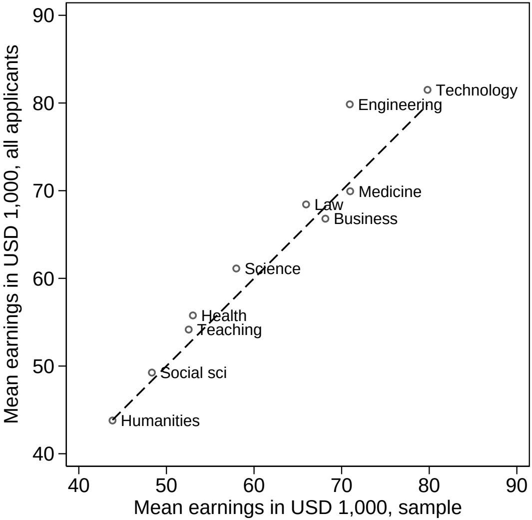
		Control function		
	(1)	(2)	(3)	(4)
Parameter	IV	Covariates	Sites	Full model
$LATE_h$	0.247	0.261	0.190	0.214
	(0.031)	(0.032)	(0.076)	(0.042)
$LATE_{nh}$		0.386	0.341	0.370
		(0.143)	(0.219)	(0.088)
$LATE_{ch}$		0.023	–0.122	–0.093
		(0.251)	(0.469)	(0.154)

Payoffs to Field of Study: Kirkeboen et al. (2016)

- Kirkeboen, Leuven and Mogstad (QJE 2016) study the payoffs to field of study in Norway
- Substantive questions:
  - What are the payoffs to different fields of study, e.g., social science vs. engineering?
  - Do individuals sort across fields according to comparative advantage?
- Different angle on returns to institutions and selectivity we saw earlier

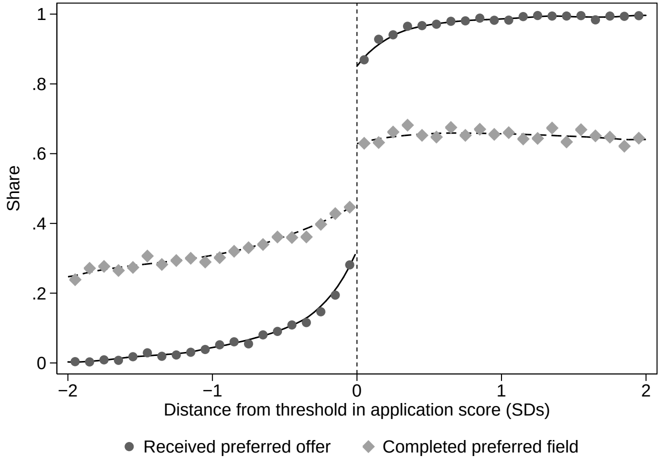
Kirkeboen et al. (2016): Context

- Context: Norwegian higher education
- Norway has a centralized admissions process
  - Apply to field/institution simultaneously (e.g. teaching at University of Oslo)
  - Rank up to 15 choices
  - Applications scored based on high school GPA, then ranked by application score
  - Then places are assigned in turn: Best applicant gets favorite choice, next best gets highest choice for which he qualifies, and so on
- Rank cutoffs generate instruments for every field



Note: This figure reports mean earnings by field for our sample of applicants and for all applicants. Earnings are measured eight years after application. The measures of earnings are regression adjusted for year of application.

Figure 2. Mean earnings by field: Sample and all applicants



Note: This figure shows the sample fraction that is offered or complete the preferred field by application score. We pool all admission cutoffs and normalize the data so that zero on the x-axis represents the admission cutoff to the preferred field. We plot unrestricted means in bins and include estimated local linear regression lines on each side of the cutoff.

Figure 3. Admission thresholds and preferred field offer and completion

Kirkeboen et al. (2016): Notation

- Consider a potential outcomes model with $J + 1$ fields of study
- Potential outcomes: $Y_i(0)$, $Y_i(1)$, ..., $Y_i(J)$
- Interested in estimating averages of treatment effects:

$$\Delta_i(j, k) = Y_i(j) - Y_i(k)$$

- Instrument is assigned field: $Z_i \in \{0, 1, ..., J\}$
- Potential treatment choices: $D_i(0)$, $D_i(1)$,...,$D_i(J)$
- Treatment and instrument indicators: $D_{ij}(z) = 1\{D_i(z) = j\}$, $Z_{ij} = 1\{Z_i = j\}$

Kirkeboen et al. (2016): Framework

- Assumptions:

$$(Y_i(0), ..., Y_i(J), D_i(0), ..., D_i(J)) \perp\!\!\!\perp Z_i$$

$$D_{ij}(j) \geq D_{ij}(k) \ , \ \forall i, j, k$$

- The second assumption says that moving the instrument from k to j makes everyone more likely to choose j - extension of monotonicity
- Think of this as an "encouragement design" where Z_i is offered field and D_i is enrolled field

Kirkeboen et al. (2016): Empirical Strategy

- Tempting to estimate a 2SLS model with J endogenous variables:

$$Y_i = \alpha + \sum_{j=1}^{J} \beta_j D_{ij} + \epsilon_i$$

$$D_{ij} = \lambda_j + \sum_{k=1}^{J} \pi_{jk} Z_{ik} + \eta_{ij}$$

- It turns out that β_j is not generally interpretable as an average treatment effect under our assumptions – LATE theorem doesn't generalize to multiple endogenous variables
- Issue: Moving Z_i from k to j makes everyone more likely to choose j but may shift people across all other pairs of fields, creating $(J+1)J$ compliance groups
- Need restrictions on effect heterogeneity or substitution patterns to interpret β_j as a LATE

Kirkeboen et al. (2016): Empirical Strategy

- Convenient feature of centralized assignment: the fallback field is known for everyone
- Conditional on fallback (below-threshold) field k, can identify LATE for compliers who switch from k to j at the threshold:

$$\beta_{jk} = E\left[Y_i(j) - Y_i(k) | D_i(j) = j, D_i(k) = k,\ j \text{ ranked above } k\right]$$

- Similarly, conditional on fallback j, can identify LATE for compliers who switch from j to k:

$$\beta_{kj} = E\left[Y_i(k) - Y_i(j) | D_i(k) = k, D_i(j) = j,\ k \text{ ranked above } j\right]$$

- A constant-effects model implies $\beta_{jk} = -\beta_{kj}$. What about a Roy model?

TABLE IV

2SLS ESTIMATES OF THE PAYOFFS TO FIELD OF STUDY (USD 1,000)

	Next Best Alternative (k)								
	Humanities	Soc Science	Teaching	Health	Science	Engineering	Technology	Business	Law
Completed field (j)									
Humanities		21.4*	−4.7	−22.9*	5.0	−38.5**	6.9	−42.2**	−156.3
		(11.0)	(9.8)	(12.1)	(11.9)	(14.7)	(48.3)	(10.6)	(437.3)
Social Science	18.7**		9.8	−10.8	55.5**	−55.4**	−110.4	−28.4**	−76.1
	(6.7)		(11.6)	(13.0)	(21.5)	(20.6)	(103.0)	(10.7)	(86.4)
Teaching	22.3**	31.4**		1.8	23.5**	−33.9**	−35.3	−21.1**	22.8
	(5.0)	(7.9)		(6.6)	(9.5)	(12.5)	(37.1)	(7.1)	(127.9)
Health	18.8**	30.7**	7.7**		28.9**	−27.9**	−43.4**	−17.4**	−55.2
	(6.3)	(7.6)	(2.8)		(7.6)	(10.4)	(20.8)	(4.0)	(97.7)
Science	53.7**	69.6**	38.6**	29.6**		−2.2	16.8	−4.9	148.3
	(18.4)	(22.4)	(14.2)	(11.5)		(14.6)	(18.1)	(10.5)	(276.2)
Engineering	59.8	−5.5	75.2**	0.2	52.4**		−46.0	−13.0	−57.7
	(50.6)	(58.2)	(37.5)	(16.4)	(21.0)		(43.9)	(23.7)	(166.6)
Technology	41.9**	58.7**	22.1*	32.5**	68.1**	−5.6		7.0	−53.1
	(10.8)	(10.1)	(12.4)	(10.1)	(9.6)	(12.0)		(9.5)	(147.5)

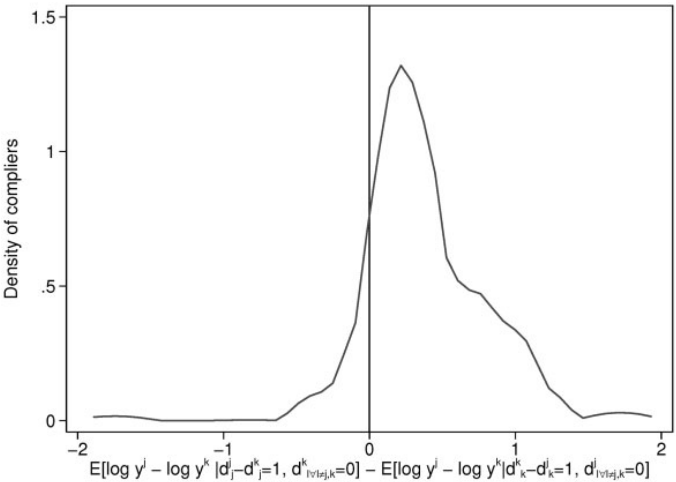


FIGURE XII

Testable Implication of Sorting Based on Comparative Advantage

This figure graphs the distribution of the differences in relative payoffs to field j versus k between individuals whose preferred field is j and next-best alternative is k, and those with the reverse ranking. To construct this graph, we use the complier-weighted distribution of estimates in Online Appendix Table B.VI.

Selection into Charter School Lotteries: Walters (2018)

- Selection models can be useful for thinking about external validity: who participates in the experiment?
- Walters (JPE 2018) studies this question in the context of charter schools
  - Publicly funded schools operating outside of traditional public school districts
  - When oversubscribed, admit applicants by random lottery
  - Studies based on lottery applicants show that urban charters boost achievement (Abdulkadiroglu et al., 2011; Dobbie and Fryer, 2011)
  - But lottery applicants are a small, highly-selected population. Would charter expansion produce gains for broader groups of students?
- Walters (2018) models selection into charter application with a dynamic generalized Roy model

Walters (2018): Setting

- Setting: charter schools in Boston
- Boston charters employ "No Excuses" practices
  - Extended instructional time, strict discipline, high expectations
  - Earlier lottery studies demonstrate large improvements in achievement for applicants (Abdulkadiroglu et al., 2011)
  - Continuing controversy over charter expansion
- Decentralized application process, separate for every school
  - Students must take steps outside of normal enrollment process
  - May apply to as many charters as desired, or none

Walters (2018): Selection Process

- Three stage selection process:

  1. Students decide whether to apply to charter schools, $A_i \in \{0, 1\}$
  2. Charters randomize offers among applicants, $Z_i \in \{0, 1\}$
  3. Applicants with $Z_i = 1$ decide whether to attend charter; students with $Z_i = 0$ remain in traditional public school

- Once enrolled in a school, students earn test scores Y_i

- Randomization at stage 2 makes Z_i a good instrument for charter attendance in the population with $A_i = 1$

TABLE 1
Descriptive Statistics for Boston Middle School Students

	All Boston Students		Charter Applicants	
	Mean (1)	Standard Deviation (2)	Mean (3)	Standard Deviation (4)
	A. Charter School Applications and Attendance			
Applied to charter school	.175	.380	1.000	.000
Applied to more than one charter	.046	.210	.265	.442
Received charter offer	.125	.331	.718	.450
Attended charter school	.112	.316	.600	.490
	B. Student Characteristics			
Female	.492	.500	.490	.500
Black	.460	.498	.518	.500
Hispanic	.398	.490	.317	.465
Subsidized lunch	.821	.383	.723	.448
Special education	.226	.418	.170	.376
Limited English proficiency	.212	.409	.136	.343
4th-grade math score	−.520	1.070	−.314	.990
4th-grade reading score	−.636	1.137	−.413	1.036
	C. Nearby Schools			
Miles to closest charter school	2.105	1.168	1.859	1.087
Miles to closest district school	.512	.339	.580	.372
Value-added of closest district school	.032	.154	.022	.167
Observations	9,156		1,601	

TABLE 3
TWO-STAGE LEAST SQUARES ESTIMATES OF CHARTER SCHOOL EFFECTS

		SECOND STAGE	
INSTRUMENT	FIRST STAGE (1)	Math Scores (2)	Reading Scores (3)
Lottery offer	.641	.553	.492
	(.025)	(.087)	(.092)
Observations		1,601	
Differential distance	−.026	.453	.380
	(.002)	(.212)	(.217)
Observations		9,156	

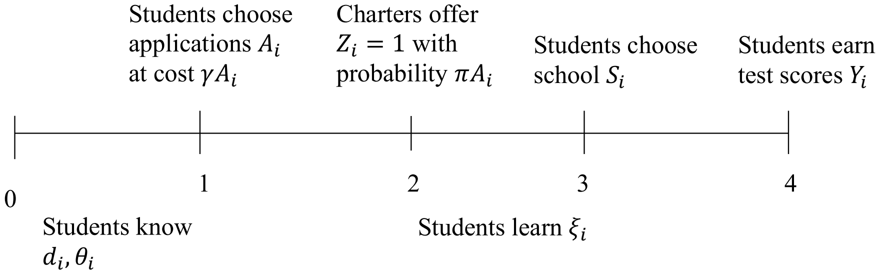
NOTE.—This table reports 2SLS estimates of the effects of charter school attendance on eighth-grade test scores. The endogenous variable is an indicator equal to one if a student attended a charter school at any time prior to the test. The first row instruments for charter attendance using a lottery offer indicator, and the second row instruments for charter attendance using distance to the closest charter school minus distance to the closest district school. Column 1 reports first-stage impacts of the instruments on charter school attendance, and cols. 2 and 3 report second-stage effects on math and reading scores. The lottery sample is restricted to charter school applicants, while the distance sample includes all Boston students. The lottery models control for lottery portfolio indicators. The distance models control for sex, race, subsidized lunch, special education, limited English proficiency, the value-added of the closest traditional public school, and fourth-grade math and reading scores.

Walters (2018): Single-Charter Selection Model

- Student i's utility from attending a charter school ($S_i = 1$):

$$U_{i1} = \mu(d_i) + v_i$$

- d_i is distance to charter
- v_i: Unobserved taste for charter schools
    - Decompose into $v_i = \theta_i + \xi_i$, with ξ_i learned after application
- Utility of attending traditional public school normalized to $U_{i0} = 0$
- Charter applicants pay utility cost γ
- Generalized Roy model: v_i may be related to potential outcomes in charter and traditional schools, $Y_i(1)$ and $Y_i(0)$


Students choose applications A_i at cost γA_i
Charters offer $Z_i = 1$ with probability πA_i
Students choose school S_i
Students earn test scores Y_i
0
1
2
3
4
Students know d_i, θ_i
Students learn ξ_i

Walters (2018): Enrollment Stage

- Solve the model by backward induction
- At the enrollment stage, students choose schools to maximize utility:

$$S_i = \arg \max_{j \in \mathcal{C}(Z_i)} U_{ij}$$

- Choice set depends on whether the student has received a charter offer:

$$\mathcal{C}(Z_i) = \{0\} \cup \{Z_i\}$$

- Before learning ξ_i, expected utility is:

$$w(Z_i | d_i, \theta_i) = E\left[\max_{j \in \mathcal{C}(Z_i)} U_{ij} | d_i, \theta_i\right]$$

$$= \begin{cases} 0, & Z_i = 0 \\ E\left[\max\left\{\mu(d_i) + \theta_i + \xi_i, 0\right\} | d_i, \theta_i\right], & Z_i = 1 \end{cases}$$

Walters (2018): Emax

- Expected utility $w(Z_i|d_i, \theta_i)$ is increasing in Z_i because a charter offer provides an option value at the school enrollment stage
- Example of *Emax*, key concept in dynamic discrete choice
  - Students can reoptimize in response to new information, so $E[\max U_{ij}] \geq \max E[U_{ij}]$
  - Students plan ahead, knowing they will later make an Emax decision
- If $\xi_i \sim$ *logistic*, we have

$$w(Z_i|d_i, \theta_i) = \log\left(1 + Z_i \times [\mu(d_i) + \theta_i]\right)$$

$$\Pr\left[S_i = 1|Z_i, d_i, \theta_i\right] = \frac{Z_i \times \exp(\mu(d_i) + \theta_i)}{1 + Z_i \times \exp(\mu(d_i) + \theta_i)}$$

Walters (2018): Lottery Stage

- Random assignment makes the lottery stage simple
- The probability that student i receives a lottery offer is

$$\Pr[Z_i = 1|A_i] = \pi A_i$$

- Applicants receive offers with probability π, non-applicants do not receive offers

Walters (2018): Application Stage

- Forward-looking students choose applications to maximize expected utility:

$$A_i = 1\left\{\pi w(1, d_i, \theta_i) - \gamma > 0\right\}$$

- With logistic ξ_i, the application rule is:

$$A_i = 1\left\{\theta_i > \exp\left(\gamma/\pi\right) - 1 - \mu(d_i)\right\}$$

- If $\theta_i \sim N(0, \sigma_\theta^2)$, application stage becomes a probit:

$$\Pr\left[A_i = 1 | d_i\right] = \Phi\left(\frac{\mu(d_i) + 1 - \exp(\gamma/\pi)}{\sigma_\theta}\right)$$

Walters (2018): Interpreting IV

- Lottery compliers apply to charter schools, then accept offers if admitted
- LATE for lottery compliers conditional on distance is:

$$LATE(d_i) = E\left[Y_i(1) - Y_i(0) | \theta_i > \exp\left(\gamma/\pi\right) - 1 - \mu(d_i), \mu(d_i) + \theta_i + \xi_i > 0\right]$$

- If distance d_i is independent of potential outcomes, we can use variation in LATEs by distance to understand the nature of selection (requires exclusion restriction)
- Students who apply from far away are more selected than students who apply from close by
  - $LATE(d_i)$ traces out the relationship between unobserved preferences and treatment effects
  - Informative about external validity

Walters (2018): Choice Model Estimation

- In practice, Walters (2018) takes a parametric approach to estimation
- Likelihood of student i's choices:

$$\mathcal{L}(A_i, Z_i, S_i|d_i) = \int \Phi\left(\frac{\mu(d_i)-\exp(\gamma/\pi)-1}{\sigma_\theta}\right)^{A_i}\left[1-\Phi\left(\frac{\mu(d_i)-\exp(\gamma/\pi)-1}{\sigma_\theta}\right)\right]^{1-A_i}$$

$$\times \pi^{A_i Z_i}(1-\pi)^{A_i(1-Z_i)}$$

$$\times\left[\frac{\exp(\mu(d_i)+\theta_i)}{1+\exp(\mu(d_i)+\theta_i)}\right]^{A_i Z_i S_i}\left[\frac{1}{1+\exp(\mu(d_i)+\theta_i)}\right]^{A_i Z_i(1-S_i)} dF(\theta_i)$$

- Estimate parameters by simulated maximum likelihood, approximating integral by drawing from the distribution of θ_i

Walters (2018): Control Function Estimation

- Suppose potential outcomes are given by:

$$E[Y_i(s)|d_i, \theta_i] = \alpha_d + \beta_d \theta_i, \;\; s \in \{0, 1\}$$

- We can estimate these parameters via the OLS regression:

$$Y_i = \alpha_0 + (\alpha_1 - \alpha_0)S_i + \beta_0 \theta^*(A_i, Z_i, S_i, d_i) + (\beta_1 - \beta_0)S_i \theta^*(A_i, Z_i, S_i, d_i) + e_i$$

- Control function $\theta^*(A_i, Z_i, S_i, d_i)$ is predicted value of unobserved taste θ_i given i's observed choices and instruments – generalization of Heckit Mills ratio

- $\alpha_1 - \alpha_0$ is the ATE, while β_1 and β_0 govern selection

TABLE 5
CHARTER SCHOOL PREFERENCE PARAMETER ESTIMATES

	Charter School Utility (1)	Disutility of Distance (2)	Application Costs: Log Fixed Cost (3)	Application Costs: Log Marginal Cost (4)
Constant/main effect	−1.099	.182	−2.098	−.664
	(.095)	(.016)	(.187)	(.016)
Female	−.046	−.018	−.006	.027
	(.097)	(.011)	(.130)	(.026)
Black	−.465	−.156	1.286	−.241
	(.152)	(.018)	(1.035)	(.047)
Hispanic	−.376	−.128	1.713	−.232
	(.164)	(.019)	(1.041)	(.051)
Subsidized lunch	−.298	−.008	.379	.091
	(.124)	(.014)	(.210)	(.032)
Special education	−.228	−.025	.098	.025
	(.137)	(.015)	(.162)	(.039)
Limited English proficiency	−.118	.024	.038	.100
	(.148)	(.014)	(.182)	(.040)
Value-added of closest district school	−1.156	−.177	.429	−.075
	(.306)	(.035)	(.392)	(.075)
4th-grade math score	.138	.007	−.028	.009
	(.070)	(.008)	(.092)	(.019)
4th-grade reading score	.161	.008	−.067	.022
	(.073)	(.008)	(.097)	(.019)
Distance squared	. . .	.001	. . .	. . .
		(.001)		

TABLE 7
SELECTION-CORRECTED ESTIMATES OF CHARTER SCHOOL EFFECTS ON EIGHTH-GRADE TEST SCORES

	Math Scores		Reading Scores	
	Public School Outcome (1)	Charter Effect (2)	Public School Outcome (3)	Charter Effect (4)
Constant/main effect	−.390	.705	−.508	.522
	(.015)	(.092)	(.016)	(.096)
Female	−.024	.060	.184	−.019
	(.015)	(.046)	(.016)	(.048)
Black	−.193	.250	−.087	.199
	(.025)	(.073)	(.026)	(.077)
Hispanic	−.100	.260	−.041	.243
	(.025)	(.078)	(.027)	(.081)
Subsidized lunch	−.128	.192	−.126	.149
	(.022)	(.056)	(.023)	(.059)
Special education	−.370	.097	−.397	.134
	(.020)	(.065)	(.021)	(.068)
Limited English proficiency	.075	−.091	.044	−.074
	(.020)	(.069)	(.021)	(.072)
Value-added of closest district school	.136	.003	.113	−.041
	(.049)	(.138)	(.051)	(.145)
4th-grade math score	.476	−.122	.165	−.043
	(.011)	(.033)	(.011)	(.035)
4th-grade reading score	.066	−.019	.366	−.078
	(.011)	(.034)	(.011)	(.036)
Charter school preference, θ_i	.058	−.096	.046	−.039
	(.016)	(.047)	(.017)	(.049)
Idiosyncratic preference, τ_{ij}	. . .	−.017	. . .	.010
		(.052)		(.055)
p-values: no selection on unobservables	.001		.051	

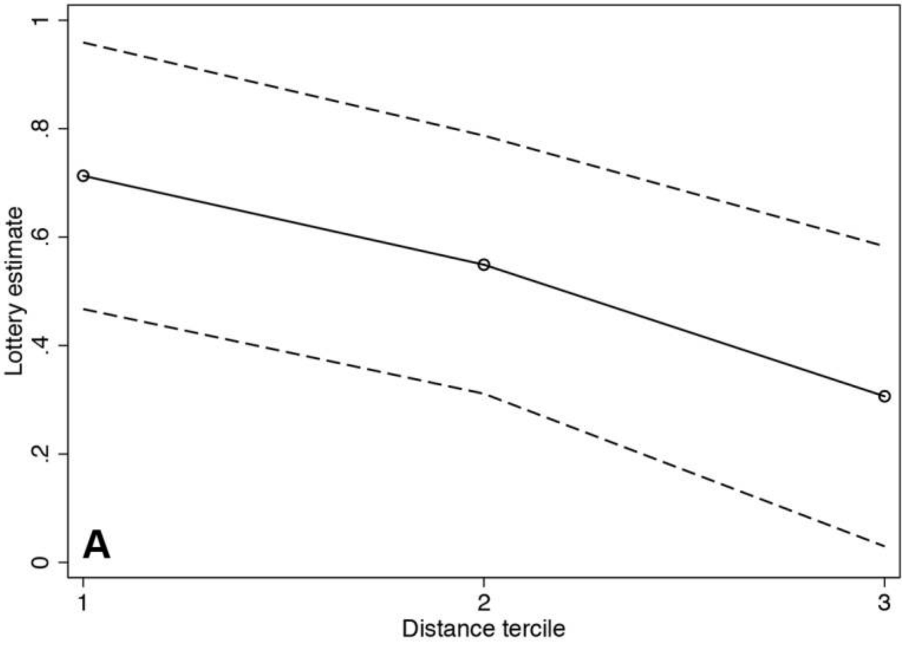
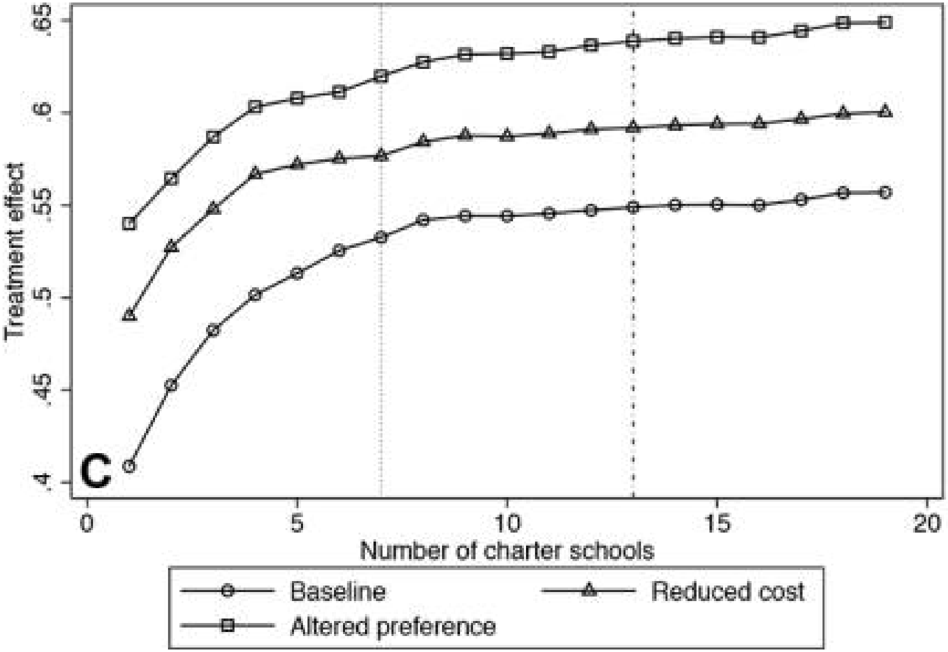

A
Lottery estimate
0
.2
.4
.6
.8
1
1
2
3
Distance tercile

TABLE 8
TEST OF RESTRICTIONS IMPLIED BY TEST SCORE MAXIMIZATION

		Math Scores		Reading Scores	
	Preference Coefficient (1)	Test Score Gain Coefficient (2)	Ratio (3)	Test Score Gain Coefficient (4)	Ratio (5)
Female	−.046	.060	−1.313	−.019	.407
Black	−.465	.250	−.538	.199	−.429
Hispanic	−.376	.260	−.691	.243	−.646
Subsidized lunch	−.298	.192	−.644	.149	−.499
Special education	−.228	.097	−.426	.134	−.588
Limited English proficiency	−.118	−.091	.773	−.074	.626
Value-added of closest district school	−1.156	.003	−.003	−.041	.036
4th-grade math score	.138	−.122	−.883	−.043	−.315
4th-grade reading score	.161	−.019	−.117	−.078	−.481
Charter school preference, θ_i	1.000	−.096	−.096	−.039	−.039
Idiosyncratic preference, τ_{ij}	1.000	−.017	−.017	.010	.010
p-values: test score maximization			<.001		<.001


C
Treatment effect
.4
.45
.5
.55
.6
.65
0
5
10
15
20
Number of charter schools
Baseline
Reduced cost
Altered preference

References

- Abadie, A. (2003). "Semiparametric instrumental variables estimation of treatment response models." *Journal of Econometrics* 113.
- Abdulkadiroglu, A., Angrist, J., Dynarski, S., Kane, T., and Pathak, P. (2011). "Accountability and flexibility in public schools: evidence from Boston's charters and pilots." *Quarterly Journal of Economics* 126.
- Brinch, C., Mogstad, M., and Wiswall, M. (2017). "Beyond LATE with a discrete instrument." *Journal of Political Economy* 125(4).
- Carneiro, P., Heckman, J., and Vytlacil, E. (2010). "Evaluating marginal policy changes and the average effect of treatment for individuals at the margin." *Econometrica* 78(1).
- Carneiro, P., and Lee, S. (2009). "Estimating distributions of potential outcomes using local instrumental variables with an application to changes in college enrollment and wage inequality." *Journal of Econometrics* 149(2).
- Dobbie, W., and Fryer, R. (2011). "Are high-quality schools enough to increase achievement among the poor? Evidence from the Harlem Children's Zone." *American Economic Journal: Applied Economics* 3(3).
- Eisenhauer, P., Heckman, J., and Vytlacil, E. (2015). "The generalized Roy model and the cost-benefit analysis of social programs." *Journal of Political Economy* 123.
- Heckman, J. (1974). "Shadow prices, market wages, and labor supply." *Econometrica* 42(4).

References

- Heckman, J. (1976). "The common structure of statistical models of truncation, sample selection and lmiited dependent variables and a simple estimator for such models." *Annals of Economic and Social Measurement* 5(4).
- Heckman, J. (1979). "Sample selection bias as a specification error." *Econometrica* 47(1).
- Heckman, J., Urzua, S., and Vytlacil, E. (2006). "Understanding instrumental variables in models with essential heterogeneity." *Review of Economics and Statistics* 88(3).
- Heckman, J., and Vytlacil, E. (1999). "Local instrumental variables and latent variable models for identifying and bounding treatment effects." *Proceedings of the National Academy of Sciences* 96(8).
- Heckman, J., and Vytlacil, E. (2005). "Structural equations, treatment effects, and econometric policy evaluation." *Econometrica* 73(3).
- Imbens, G., and Angrist, J. (1994). "Identification and estimation of local average treatment effects." *Econometrica* 62(2).
- Imbens, G., and Rubin, D. (1997). "Estimating outcome distributions for compliers in instrumental variables models." *Review of Economic Studies* 64(4).
- Kirkeboen, L., Leuven, E., and Mogstad, M. (2016). "Field of study, earnings, and self-selection." *Quarterly Journal of Economics* 131(3).

References

- Kline, P., and Walters, C. (2016). "Evaluating public programs with close substitutes: the case of Head Start." *Quarterly Journal of Economics* 131(4).
- Kline, P., and Walters, C. (2019). "On Heckits, LATE, and numerical equivalence." *Econometrica* 87(2).
- Olsen, R. (1980). "A least squares correction for selectivity bias." *Econometrica* 48(7).
- Roy, Andrew (1951). "Some Thoughts on the Distribution of Earnings." *Oxford Economic Papers New Series* 3.
- Vytlacil, E. (2002). "Independence, monotonicity and latent index models: an equivalence result." *Econometrica* 70(1).
- Walters, C. (2018). "The demand for effective charter schools." *Journal of Political Economy* 126(6).

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture 3 - Discrimination

Chris Walters

University of California, Berkeley and NBER

Labor Market Discrimination

- This lecture covers labor market discrimination
- Large gaps in labor market outcomes across demographic groups, e.g. by race, sex, and age
  - Wages
  - Labor force participation
  - Unemployment rates
  - Occupations, job mobility, non-wage compensation
- Theories of discrimination offer explanations for why group membership *per se* might be important
- Evidence looks at effects of group membership on outcomes
- See Altonji and Blank (1999), Lang and Lehmann (2012), and Guryan and Charles (2013) for reviews

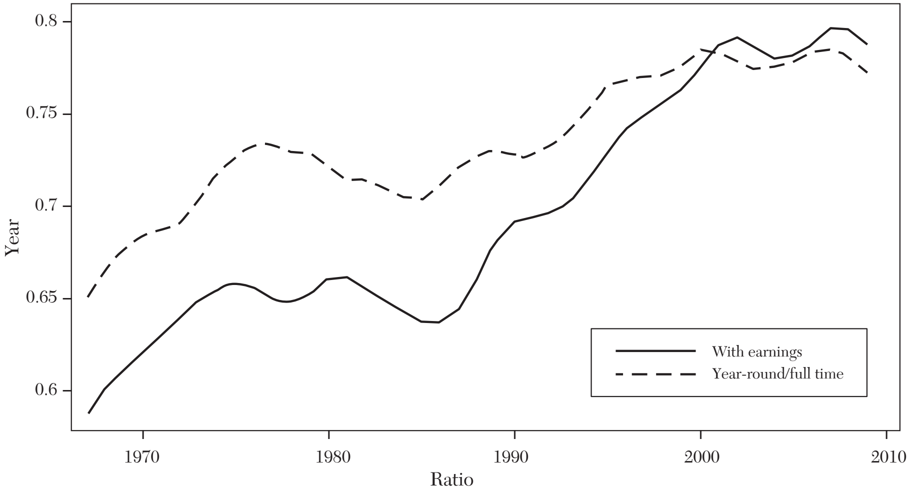
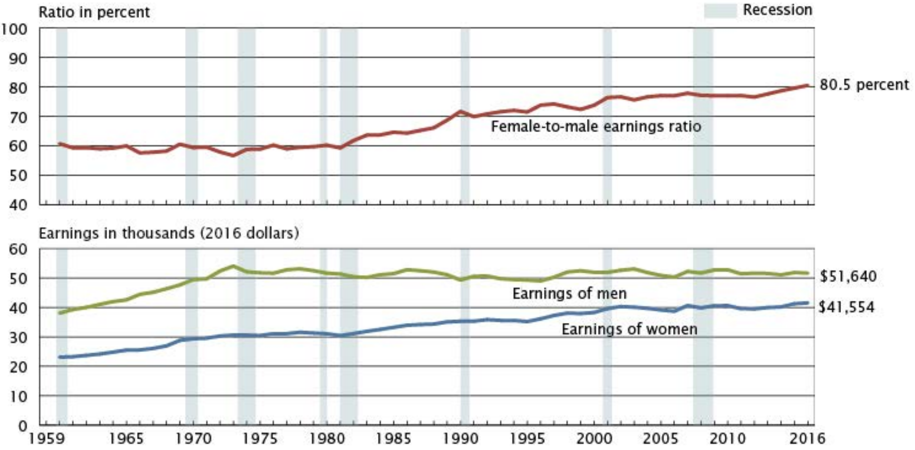


Figure 1. Ratio of Median Earnings: Black Men/White Men, 1967–2009

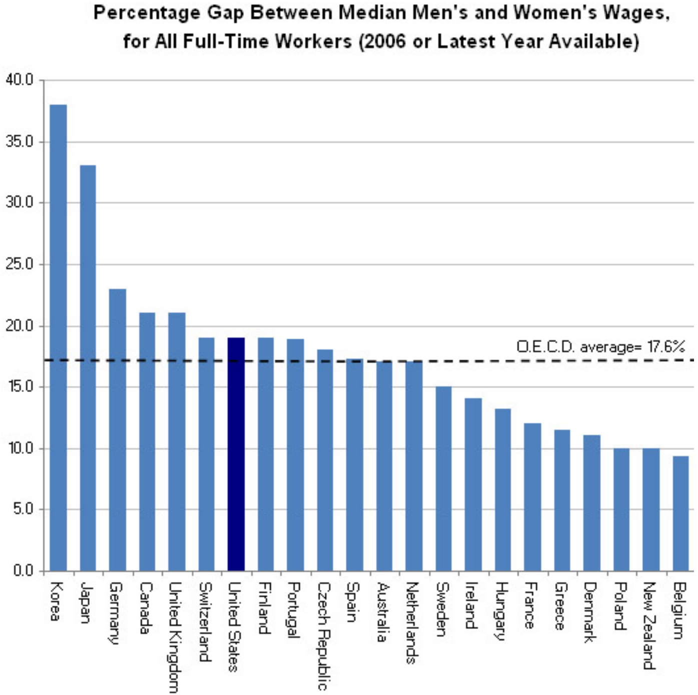
Source: Lang and Lehmann (2012)

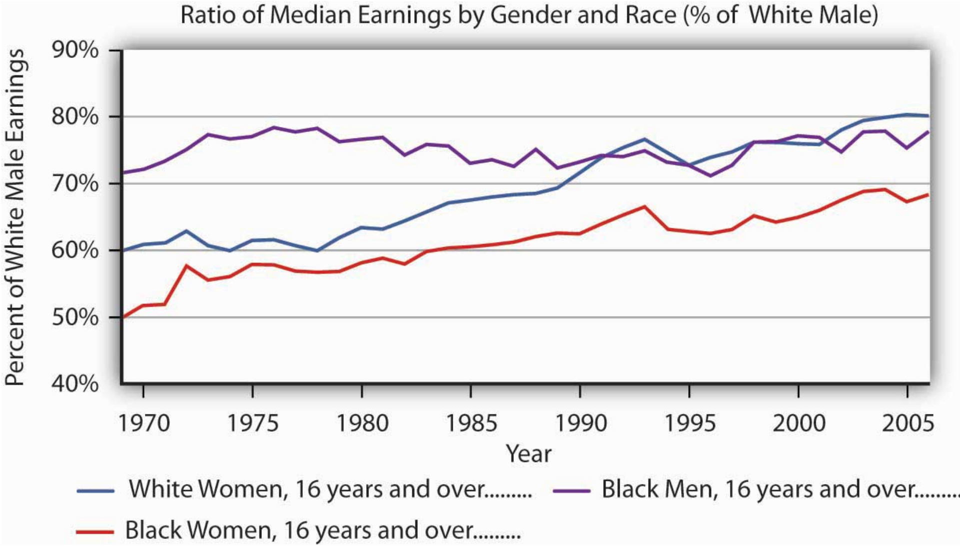
Figure 2.
Female-to-Male Earnings Ratio and Median Earnings of Full-Time, Year-Round Workers 15 Years and Older by Sex: 1960 to 2016

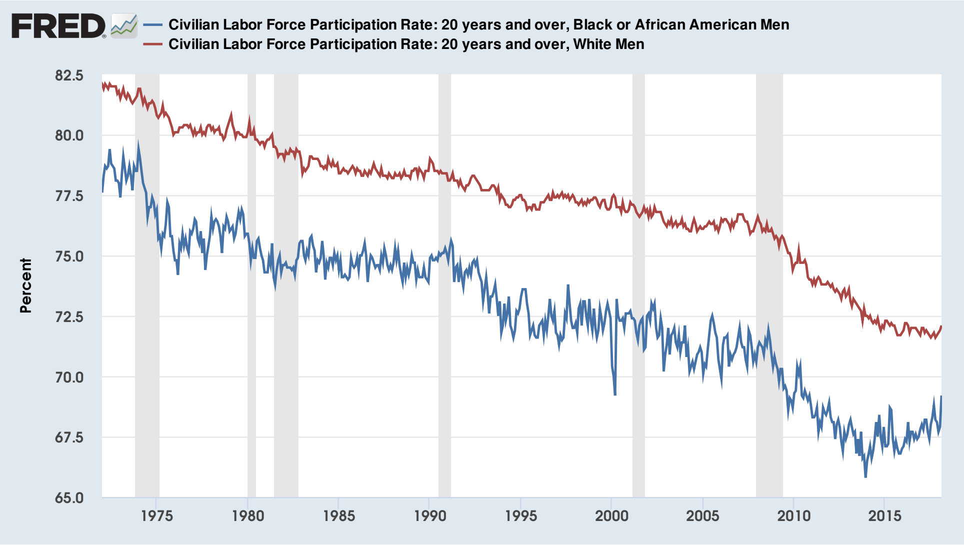


Note: The data for 2013 and beyond reflect the implementation of the redesigned income questions. The data points are placed at the midpoints of the respective years. Data on earnings of full-time, year-round workers are not readily available before 1960. For more information on recessions, see Appendix A. For information on confidentiality protection, sampling error, nonsampling error, and definitions, see <www2.census.gov/programs-surveys/cps/techdocs/cpsmar17.pdf>.

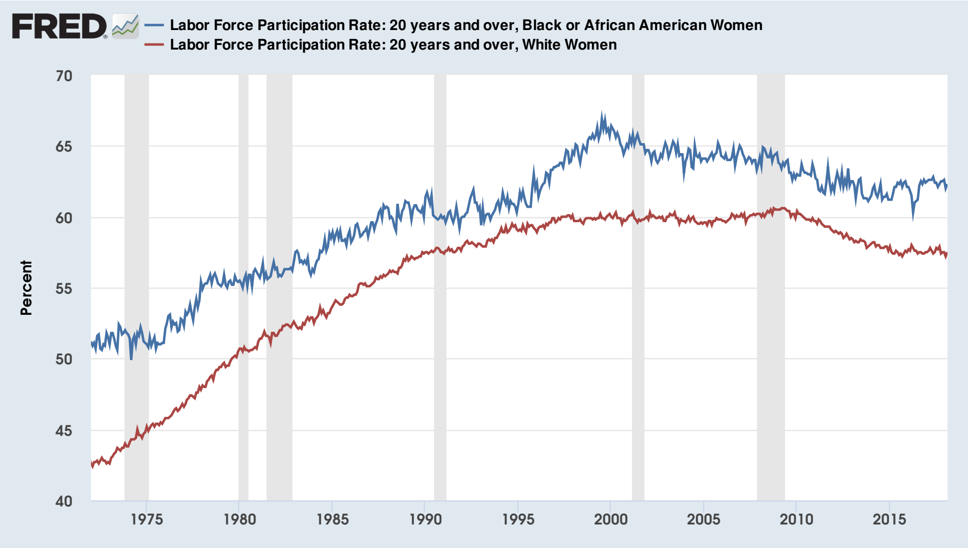
Source: U.S. Census Bureau, Current Population Survey, 1961 to 2017 Annual Social and Economic Supplements.


Percentage Gap Between Median Men's and Women's Wages,
for All Full-Time Workers (2006 or Latest Year Available)
40.0
35.0
30.0
25.0
20.0
15.0
10.0
5.0
0.0
O.E.C.D. average= 17.6%
Korea
Japan
Germany
Canada
United Kingdom
Switzerland
United States
Finland
Portugal
Czech Republic
Spain
Australia
Netherlands
Sweden
Ireland
Hungary
France
Greece
Denmark
Poland
New Zealand
Belgium

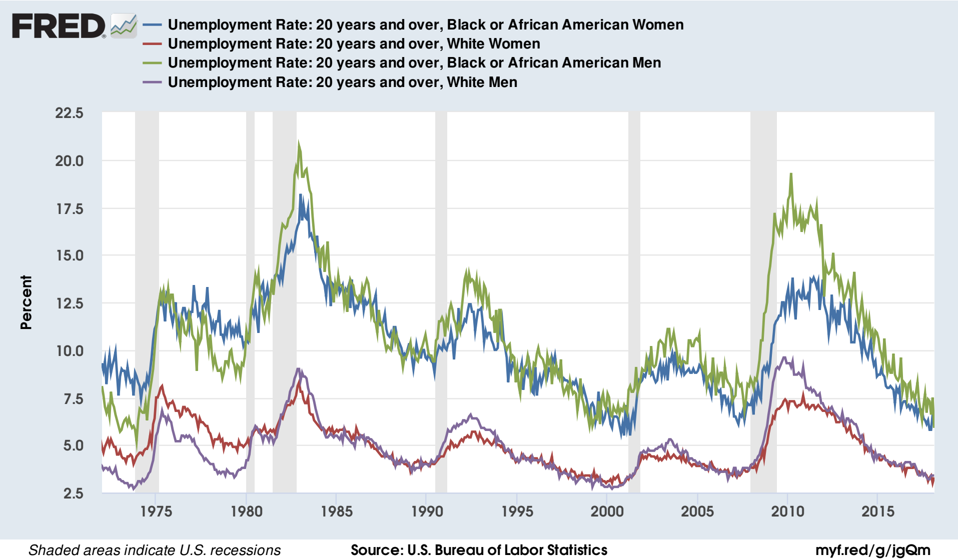

Ratio of Median Earnings by Gender and Race (% of White Male)
90%
80%
70%
60%
50%
40%
Percent of White Male Earnings
1970
1975
1980
1985
1990
1995
2000
2005
Year
White Women, 16 years and over.........
Black Men, 16 years and over.........
Black Women, 16 years and over.........



Shaded areas indicate U.S. recessions **Source: U.S. Bureau of Labor Statistics** **myf.red/g/jgQs**



Shaded areas indicate U.S. recessions **Source: U.S. Bureau of Labor Statistics** **myf.red/g/jgQw**


FRED
Unemployment Rate: 20 years and over, Black or African American Women
Unemployment Rate: 20 years and over, White Women
Unemployment Rate: 20 years and over, Black or African American Men
Unemployment Rate: 20 years and over, White Men
Percent
22.5
20.0
17.5
15.0
12.5
10.0
7.5
5.0
2.5
1975
1980
1985
1990
1995
2000
2005
2010
2015
Shaded areas indicate U.S. recessions
Source: U.S. Bureau of Labor Statistics
myf.red/g/jgQm

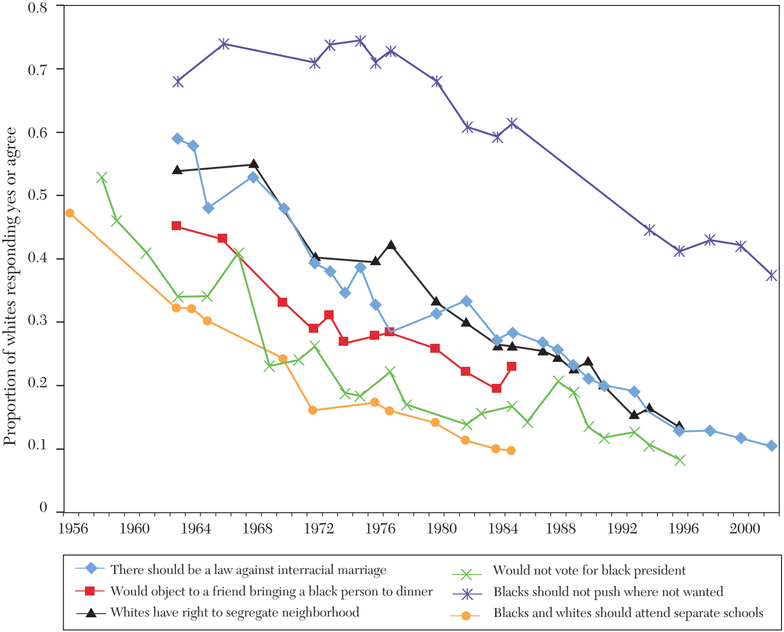


Figure 3. Trends in Prejudice Measures, 1956–2003.

Defining Discrimination

- What is discrimination? Arrow (1973):

  "The fact that different groups of workers, be they skilled or unskilled, black or white, male or female, receive different wages, invites the explanation that the different groups must differ according to some characteristic valued on the market. In standard economic theory, we think first of differences in productivity. The notion of discrimination involves the additional concept that personal characteristics of the worker unrelated to productivity are also valued in the market."

- This is a starting point, but as Altonji and Blank (1999) point out:
  - Defining "productivity" is not straightforward
  - Human capital investments (or technological changes) that affect productivity may be altered by discrimination

Theories of Discrimination

- Theories of discrimination generally fall into two broad categories:
  - **Taste-based discrimination:** Employers have prejudices that favor one group over another (Becker, 1957)
  - **Statistical discrimination:** Employers use group membership to make inferences about productivity (Aigner and Cain, 1977)
- N.B.: Both are illegal with respect to treatment of protected groups – race, color, religion, sex (including gender identity and pregnancy), national origin, age, disability, genetic information, sexual orientation, or parental status
- We will mostly focus on empirical evidence regarding the effects of protected characteristics rather than trying to distinguish between types of discrimination

Oaxaca-Blinder Decompositions

- Classic tool for measuring discrimination: the **Oaxaca-Blinder decomposition** (Oaxaca, 1973; Blinder, 1973)
- OB method decomposes a difference between groups into a component explained by observed characteristics, and a component explained by returns to characteristics
- Consider individuals in two groups, $G_i \in \{A, B\}$
- Group average outcomes are $\bar{Y}_A$ and $\bar{Y}_B$, where $\bar{Y}_g \equiv E\left[Y_i | G_i = g\right]$
- We hope to explain group differences with a vector of observed covariates X_i

Oaxaca-Blinder Decompositions

- Quantity to be explained:

$$\Delta \equiv \bar{Y}_A - \bar{Y}_B$$

- Run a separate regression for each group:
  - Group A: $Y_i = X_i'\beta_A + \epsilon_i$
  - Group B: $Y_i = X_i'\beta_B + \epsilon_i$
- X_i includes a constant
- OLS coefficient vector for group g:

$$\beta_g = E\left[X_i X_i' | G_i = g\right]^{-1} E\left[X_i Y_i | G_i = g\right]$$

Oaxaca-Blinder Decompositions

- By construction OLS fits each group's average:

$$\bar{Y}_g = \bar{X}_g' \beta_g.$$

- Therefore we can write

$$\Delta = \bar{X}_A' \beta_A - \bar{X}_B' \beta_B$$

$$= \underbrace{\left(\bar{X}_A - \bar{X}_B\right)' \beta_A}_{Explained\ \ by\ \ X's} + \underbrace{\bar{X}_B' \left(\beta_A - \beta_B\right)}_{Explained\ \ by\ \ \beta's}$$

- First term answers the question: How much more would A's make than B's if both groups were paid like A's for observables?
- Second term answers the question: How much more would A's make than B's if both groups had the B's observables?
- If X includes all characteristics relevant to productivity, second term may be attributable to discrimination

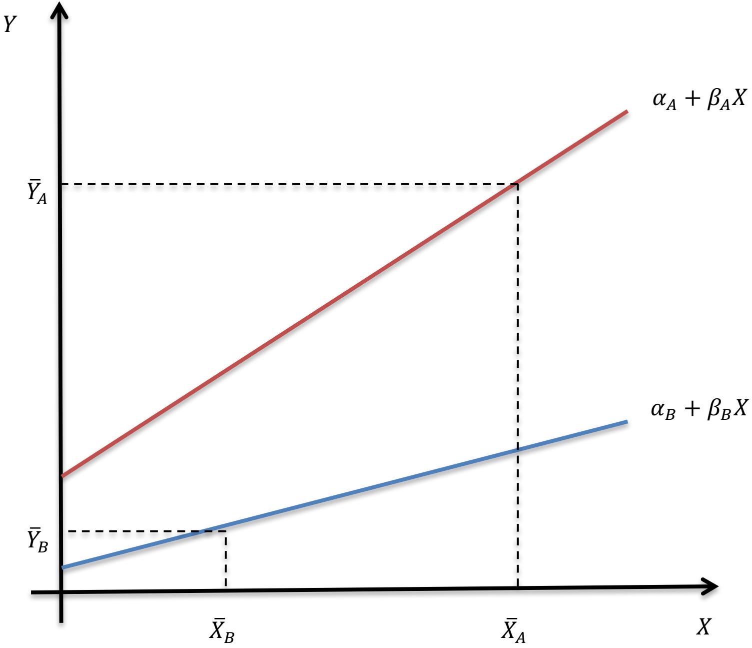
Oaxaca-Blinder Decompositions

$$\Delta = \underbrace{\left(\bar{X}_A - \bar{X}_B\right)' \beta_A}_{Explained \;\; by \;\; X's} + \underbrace{\bar{X}_B' \left(\beta_A - \beta_B\right)}_{Explained \;\; by \;\; \beta's}$$

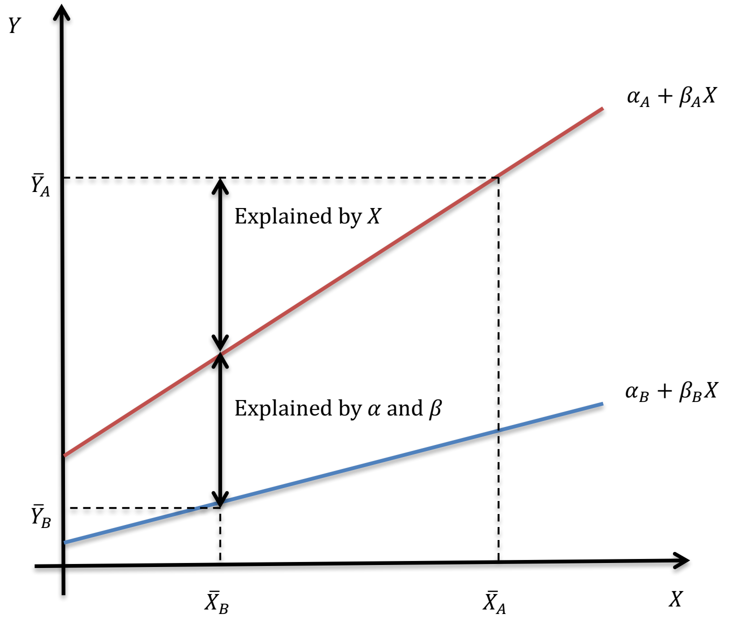
- Can also write the alternative decomposition:

$$\Delta = \underbrace{\left(\bar{X}_A - \bar{X}_B\right)' \beta_B}_{Explained \;\; by \;\; X's} + \underbrace{\bar{X}_A' \left(\beta_A - \beta_B\right)}_{Explained \;\; by \;\; \beta's}$$

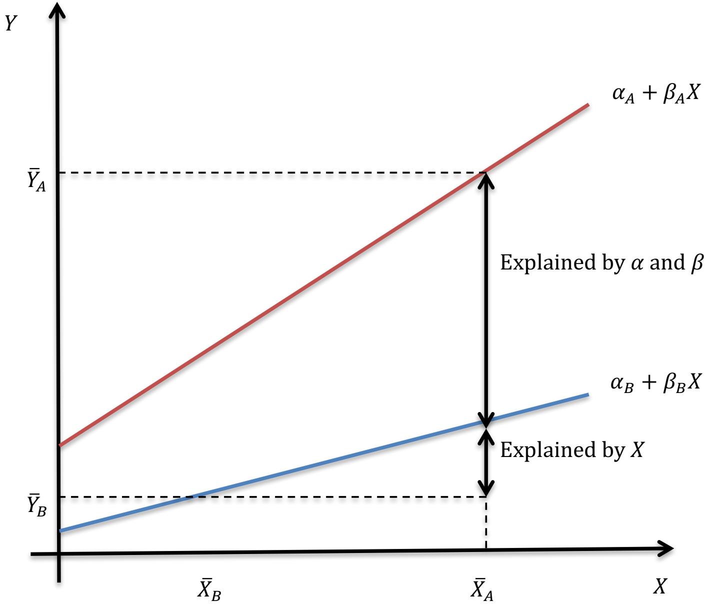
- New first term answers the question: How much more would A's make than B's if both groups were paid like B's for observables?
- New second term answers the question: How much more would A's make than B's if both groups had the A's observables?



Y
$\alpha_A + \beta_A X$
$\bar{Y}_A$
$\alpha_B + \beta_B X$
$\bar{Y}_B$
$\bar{X}_B$
$\bar{X}_A$
X



Decomposition 1: What if $B's$ had the same return to X as $A's$?



Decomposition 2: What if $A's$ had the same return to X as $B's$?

OB Decompositions and Causality

- There is a close connection between OB decompositions and our discussion of estimating treatment effects under CIA
- Let $D_i = 1\{G_i = A\}$ denote membership in group A, and re-interpret group as treatment status in a selection on observables scenario
- Let $Y_i(1)$ and $Y_i(0)$ denote i's potential outcomes, and suppose CIA holds: $(Y_i(1), Y_i(0)) \perp\!\!\!\perp D_i | X_i$
- Consider a linear model for the conditional mean of each potential outcome:

$$E\left[Y_i(d)|X_i\right] = X_i'\beta_d, \ \ d \in \{0, 1\}$$

- CIA implies β_d can be obtained by regressing Y_i on X_i in the sample with $D_i = d$

Oaxaca Treatment Effects

- Once we have the β_d's, we can use them to compute any treatment effect of interest
- Oaxaca-Blinder versions of average treatment effect parameters:

$$TOT_{OB} = E\left[Y_i|D_i = 1\right] - E\left[X_i|D_i = 1\right]' \beta_0$$

$$ATE_{OB} = E\left[X_i\right]' \left(\beta_1 - \beta_0\right)$$

$$TNT_{OB} = E\left[X_i|D_i = 0\right]' \beta_1 - E\left[Y_i|D_i = 0\right]$$

- Oaxaca uses a linear model to impute missing potential outcomes using each group's regression function

Alternative Decompositions

- Oaxaca decomposition of observed difference between treatment and control:

$$E\left[Y_i|D_i=1\right]-E\left[Y_i|D_i=0\right]=E\left[X_i|D_i=1\right]'\beta_1-E\left[X_i|D_i=0\right]'\beta_0$$

$$=\underbrace{E\left[X_i|D_i=1\right]'\left(\beta_1-\beta_0\right)}_{TOT}+\underbrace{\left(E\left[X_i|D_i=1\right]-E\left[X_i|D_i=0\right]\right)'\beta_0}_{Selection\ \ bias}$$

- We could've instead written

$$\underbrace{E\left[X_i|D_i=0\right]'\left(\beta_1-\beta_0\right)}_{TNT}+\underbrace{\left(E\left[X_i|D_i=1\right]-E\left[X_i|D_i=0\right]\right)'\beta_1}_{Selection\ \ bias}$$

- The OB decomposition is not unique for the same reason the definition of selection bias is not unique: there are multiple counterfactuals we might like to impute

Oaxaca as Reweighting

- Oaxaca-Blinder counterfactual for the treament group:

$$E[Y_i(0)|D_i = 1] = E[X_i|D_i = 1]'\beta_0$$

- Kline (2011): Can rewrite the OB counterfactual as a weighted average of control outcomes:

$$E\left[X_i|D_i = 1\right]' \beta_0 = E\left[\tilde{w}(X_i) Y_i | D_i = 0\right]$$

- Weights are

$$\tilde{w}(X_i) = X_i' E\left[X_i X_i' | D_i = 0\right]^{-1} \times E\left[X_i \frac{p\left(X_i\right)}{1 - p(X_i)} \left(\frac{1-\pi}{\pi}\right) | D_i = 0\right]$$

- Here $p(X_i) \equiv \Pr[D_i = 1|X_i]$ is the **propensity score** and $\pi = \Pr\left[D_i = 1\right]$ is the unconditional probability of treatment

- Oaxaca is a version of propensity-score reweighting

Oaxaca as Reweighting

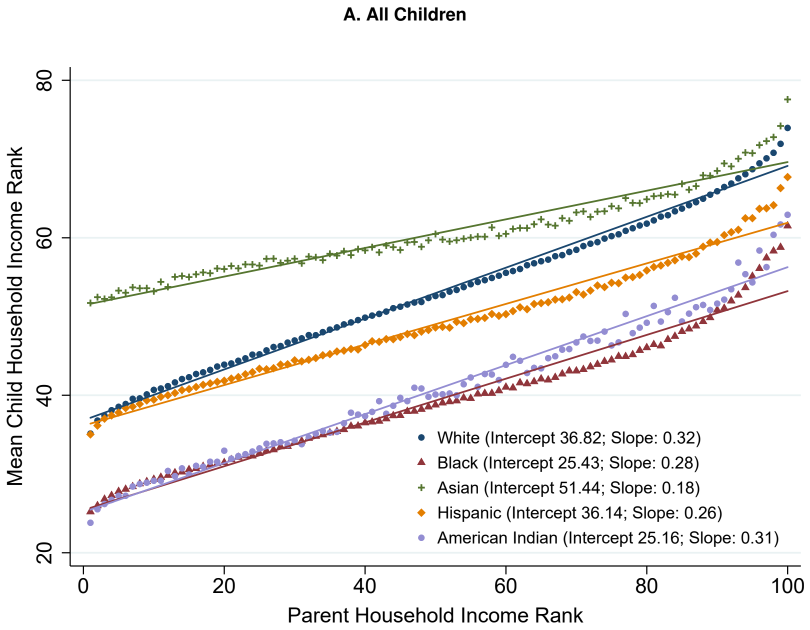
$$\tilde{w}(X_i) = X_i' E\left[X_i X_i' | D_i = 0\right]^{-1} \times E\left[X_i \frac{p(X_i)}{1 - p(X_i)} \left(\frac{1-\pi}{\pi}\right) | D_i = 0\right]$$

- Weights are fitted values from an OLS regression of the conditional odds of treatment on X_i in the control group
- Oaxaca-Blinder estimator is **doubly robust:** works if *either* $E\left[Y_i(0)|X_i\right]$ *or* $(p(X_i)/(1-p(X_i)))$ is linear in X_i
- Note that if controls are saturated linearity is guaranteed (not a restriction)
- Think of Oaxaca as another method of adjusting for observables under CIA – one that is particularly easy to implement and interpret

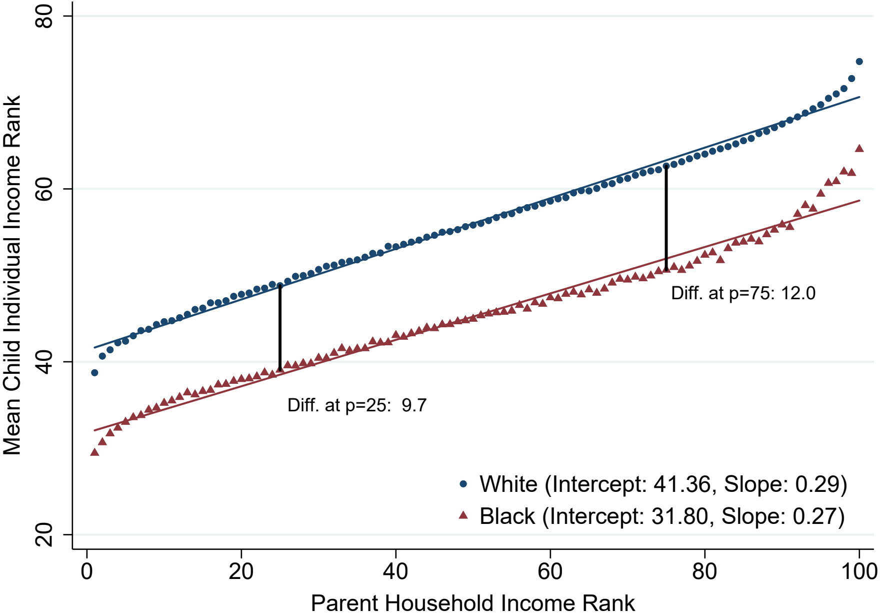
Decomposing Racial Income Gaps: Chetty et al. (2020)

- Chetty et al. (2020) perform Oaxaca-style decompositions of racial differences in income into components explained and unexplained by parent income
- Combine data from the 2000 and 2010 decennial census with federal tax returns from 1989, 1994, 1995, 1998-2015 to study variation in intergenerational mobility by race and gender
- Look at child/parent income for cohorts born 1978-1983
  - Parent income averaged over five years
  - Child income averaged over two years, at ages between 31 and 37
- Race measured in 2010 census
- Also match to American Community Survey (ACS) data – provides data on hours, wages, education, occupation

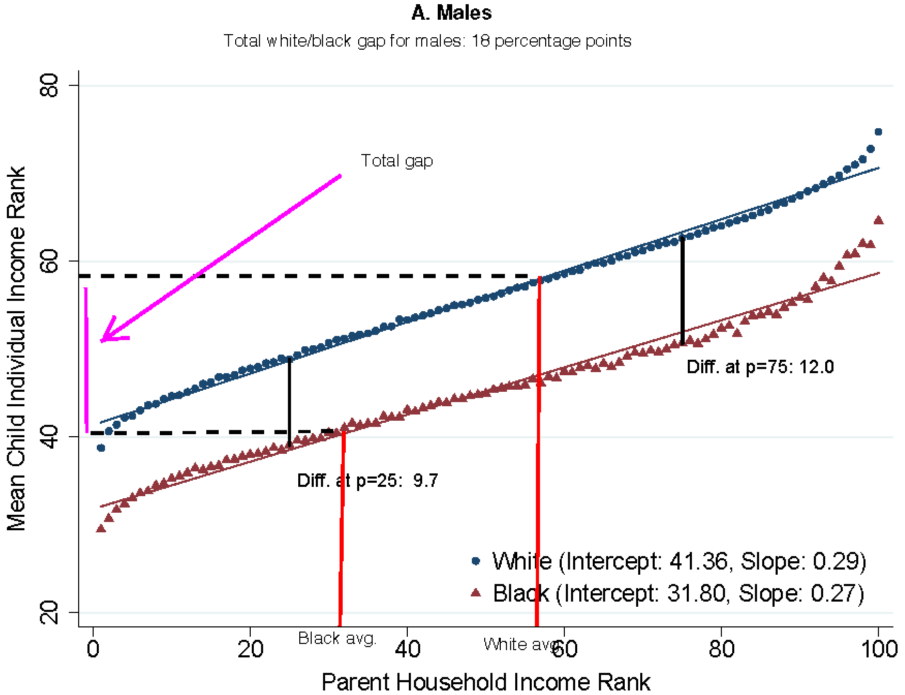
FIGURE III: Intergenerational Mobility by Race

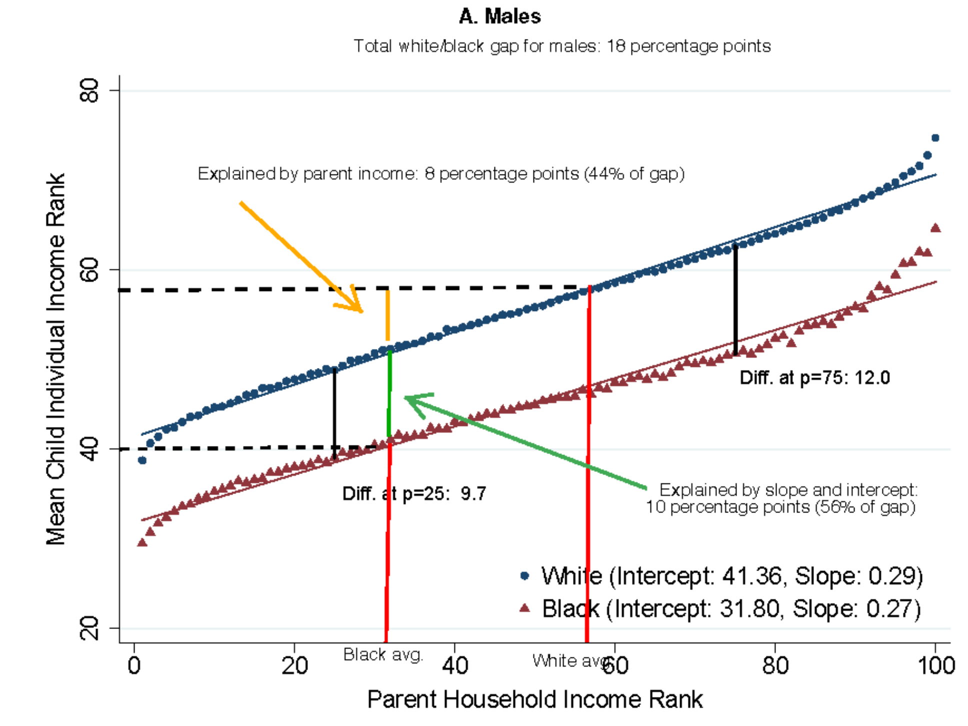


A. Males

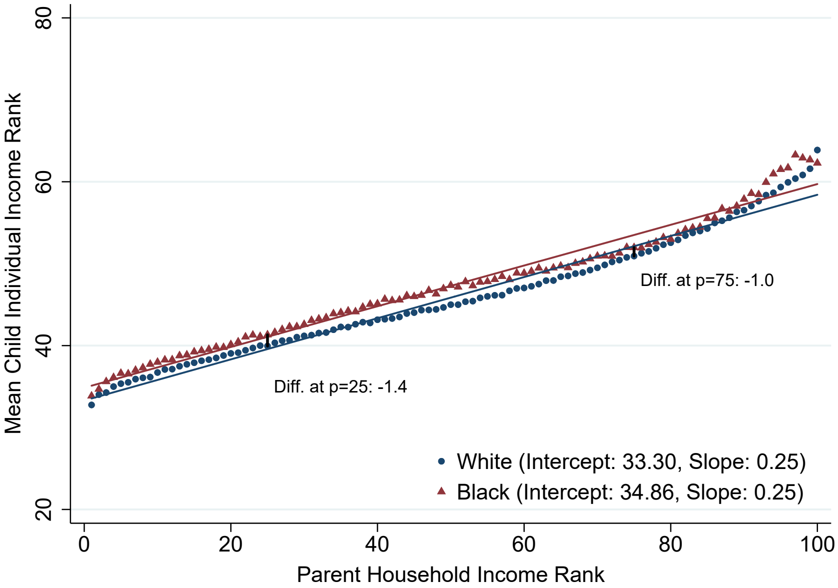


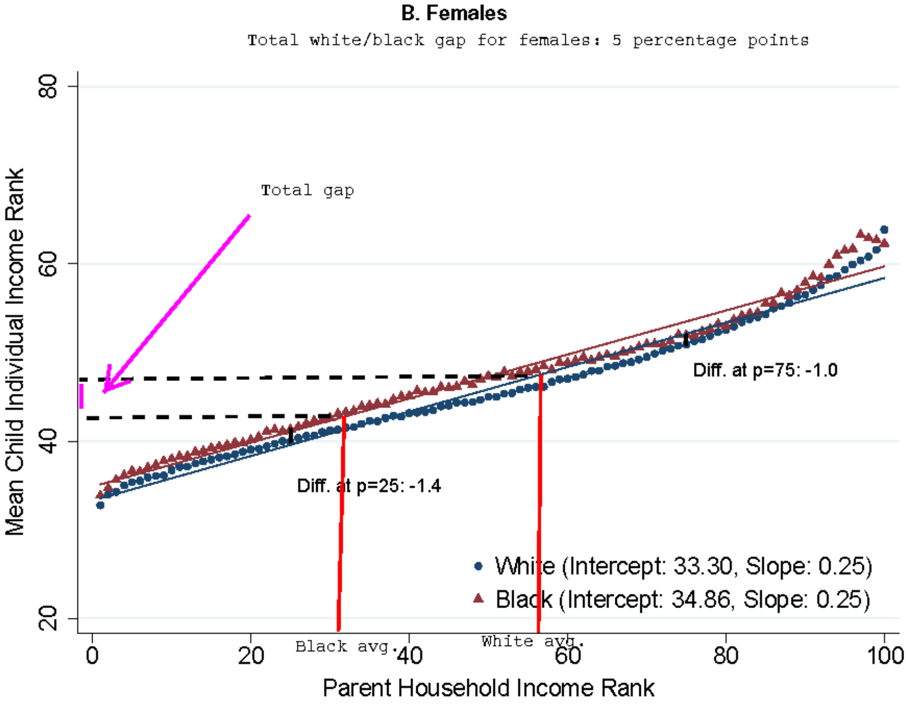
Mean Child Individual Income Rank
Parent Household Income Rank
20
40
60
80
0
20
40
60
80
100
Diff. at p=25: 9.7
Diff. at p=75: 12.0
White (Intercept: 41.36, Slope: 0.29)
Black (Intercept: 31.80, Slope: 0.27)

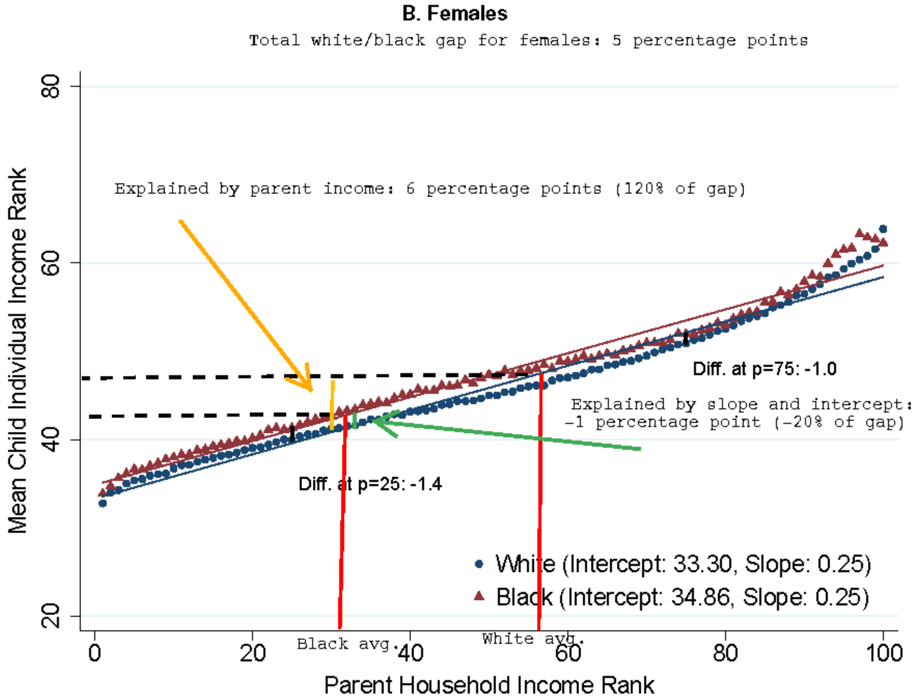

A. Males
Total white/black gap for males: 18 percentage points
Total gap
Diff. at p=75: 12.0
Diff. at p=25: 9.7
White (Intercept: 41.36, Slope: 0.29)
Black (Intercept: 31.80, Slope: 0.27)
Black avg.
White avg.
Mean Child Individual Income Rank
Parent Household Income Rank
20
40
60
80
0
20
40
60
80
100

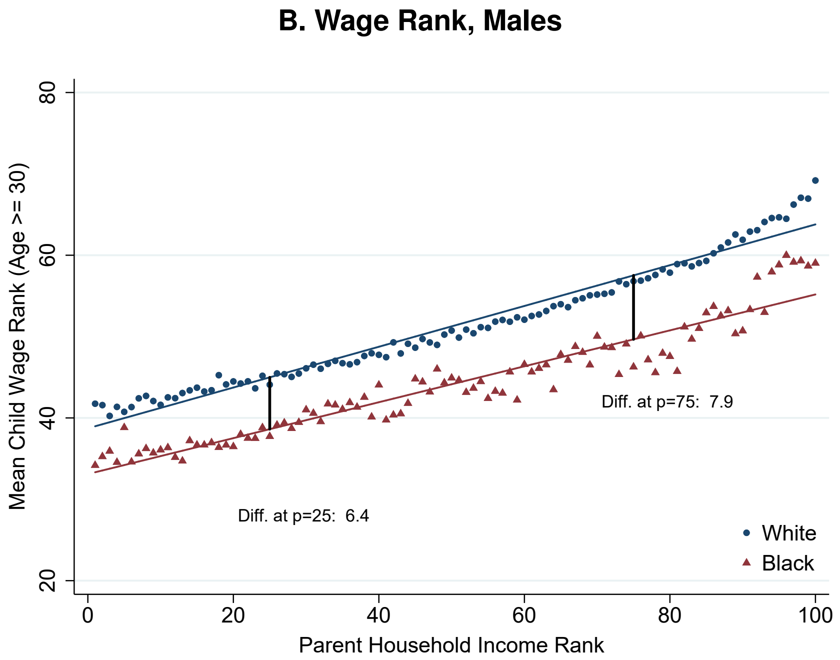

A. Males
Total white/black gap for males: 18 percentage points
Explained by parent income: 8 percentage points (44% of gap)
Diff. at p=75: 12.0
Diff. at p=25: 9.7
Explained by slope and intercept:
10 percentage points (56% of gap)
White (Intercept: 41.36, Slope: 0.29)
Black (Intercept: 31.80, Slope: 0.27)
Black avg.
White avg.
Mean Child Individual Income Rank
Parent Household Income Rank
20
40
60
80
0
20
40
60
80
100

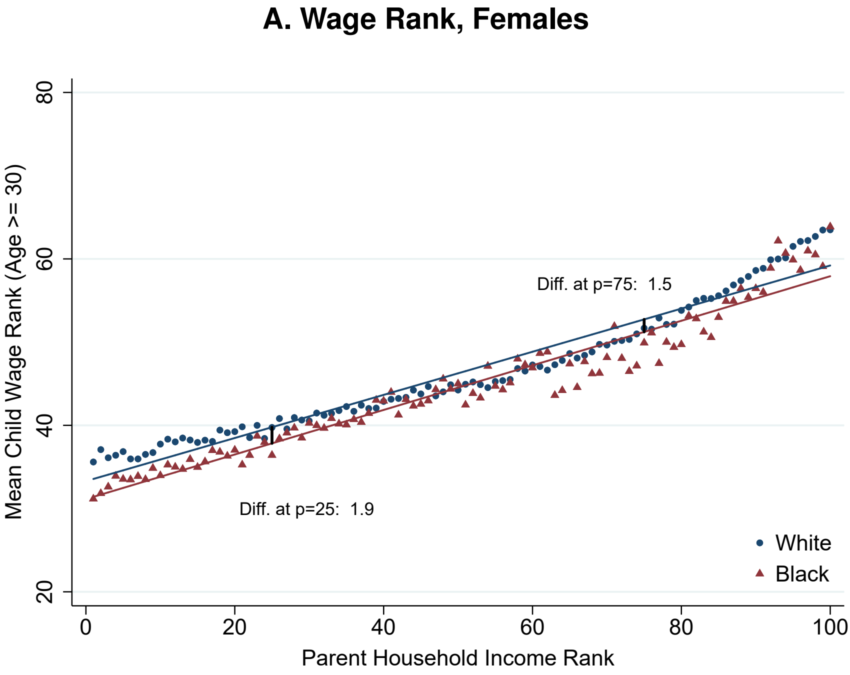
B. Females


Mean Child Individual Income Rank
80
60
40
20
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=25: -1.4
Diff. at p=75: -1.0
White (Intercept: 33.30, Slope: 0.25)
Black (Intercept: 34.86, Slope: 0.25)

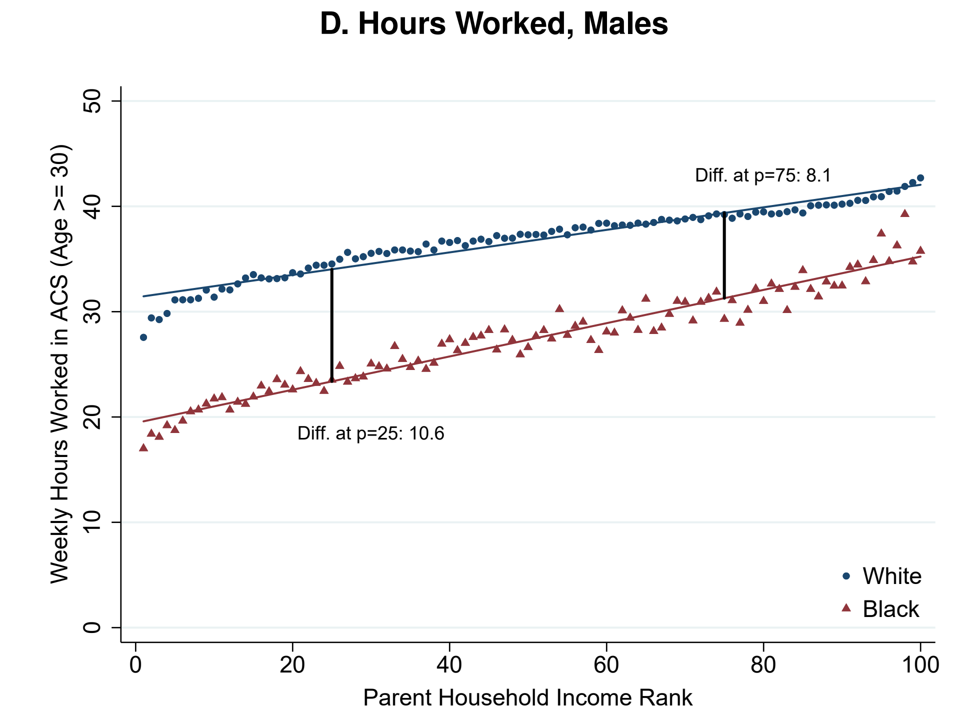

B. Females
Total white/black gap for females: 5 percentage points
Total gap
Diff. at p=75: -1.0
Diff. at p=25: -1.4
White (Intercept: 33.30, Slope: 0.25)
Black (Intercept: 34.86, Slope: 0.25)
Black avg.
White avg.
Mean Child Individual Income Rank
Parent Household Income Rank
0
20
40
60
80
100


B. Females
Total white/black gap for females: 5 percentage points
Explained by parent income: 6 percentage points (120% of gap)
Diff. at p=75: -1.0
Explained by slope and intercept: -1 percentage point (-20% of gap)
Diff. at p=25: -1.4
White (Intercept: 33.30, Slope: 0.25)
Black (Intercept: 34.86, Slope: 0.25)
Black avg.
White avg.
Mean Child Individual Income Rank
Parent Household Income Rank
20
40
60
80
0
20
40
60
80
100

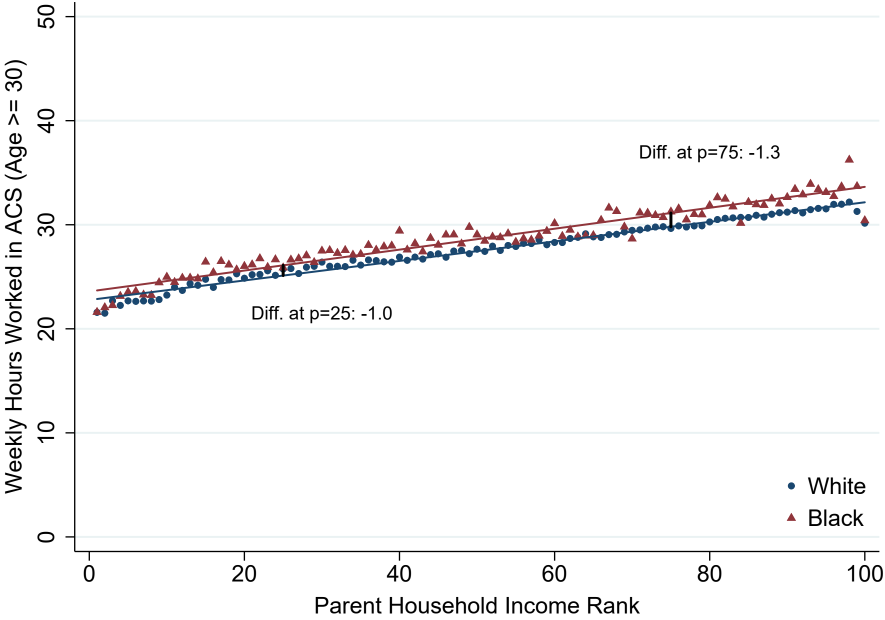

B. Wage Rank, Males
Mean Child Wage Rank (Age >= 30)
80
60
40
20
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=25: 6.4
Diff. at p=75: 7.9
White
Black

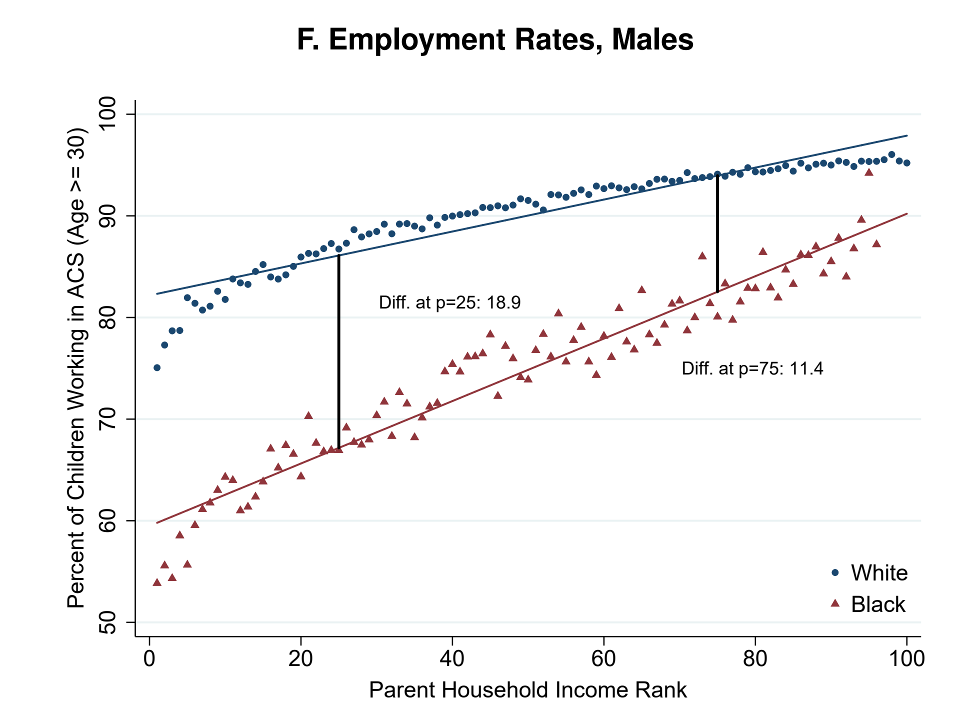


A. Wage Rank, Females
Mean Child Wage Rank (Age >= 30)
Parent Household Income Rank
Diff. at p=75: 1.5
Diff. at p=25: 1.9
White
Black
20
40
60
80
0
20
40
60
80
100

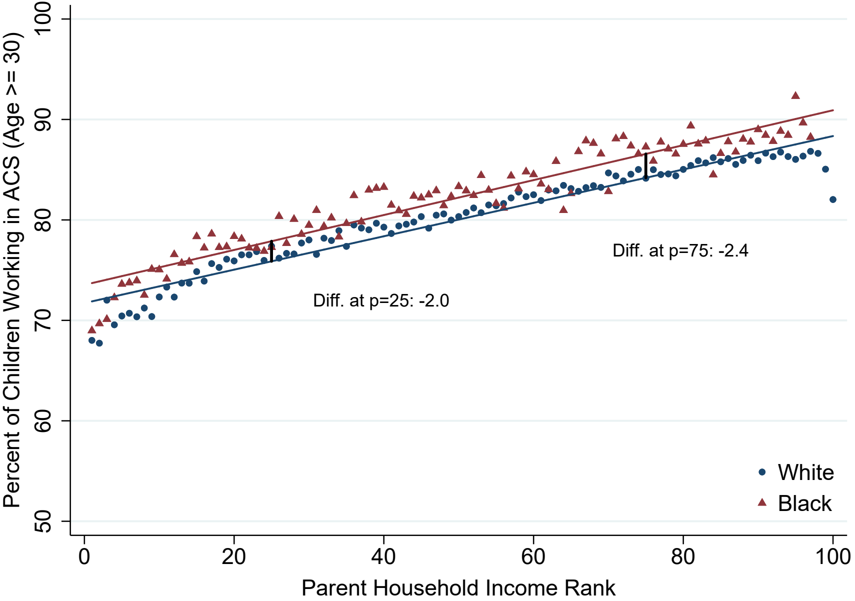

D. Hours Worked, Males
Weekly Hours Worked in ACS (Age >= 30)
0
10
20
30
40
50
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=75: 8.1
Diff. at p=25: 10.6
White
Black

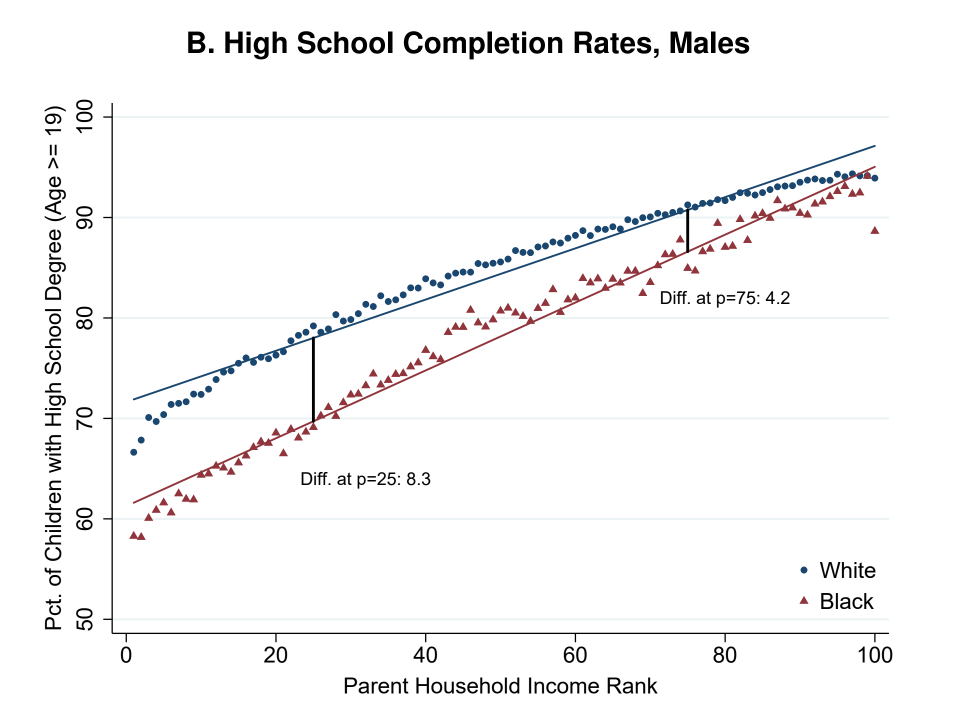
C. Hours Worked, Females

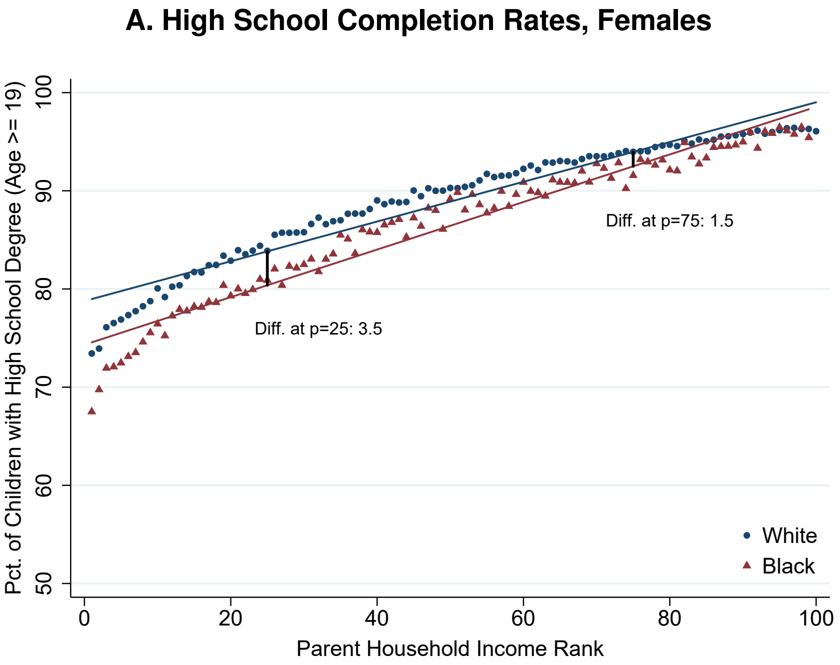


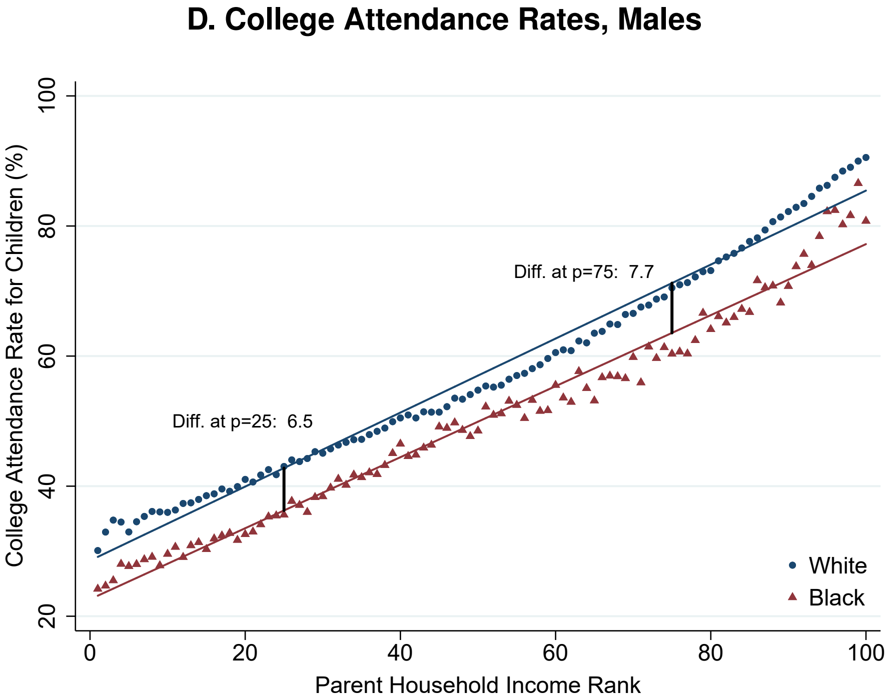

F. Employment Rates, Males
Percent of Children Working in ACS (Age >= 30)
100
90
80
70
60
50
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=25: 18.9
Diff. at p=75: 11.4
White
Black

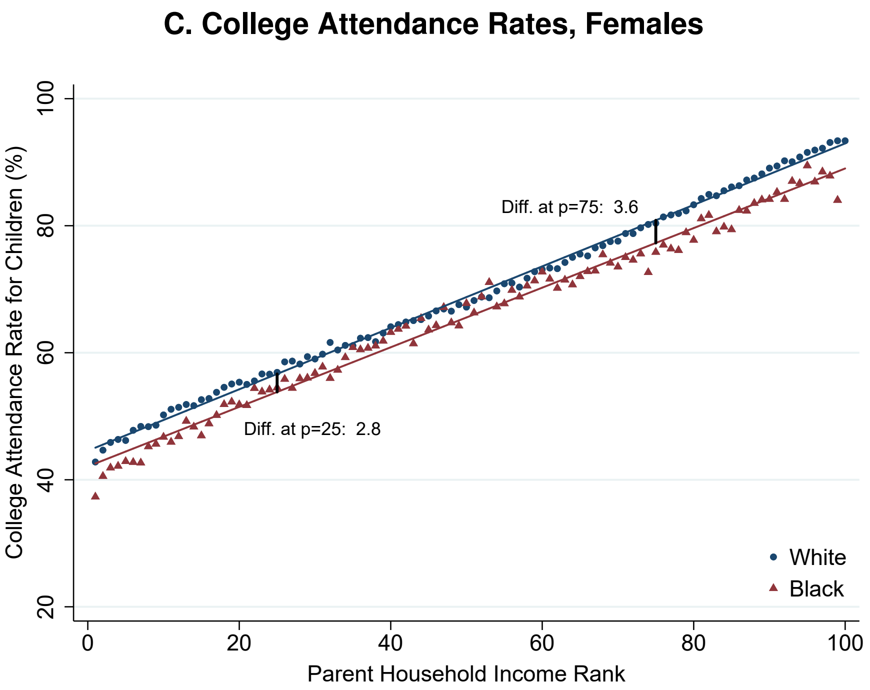
E. Employment Rates, Females

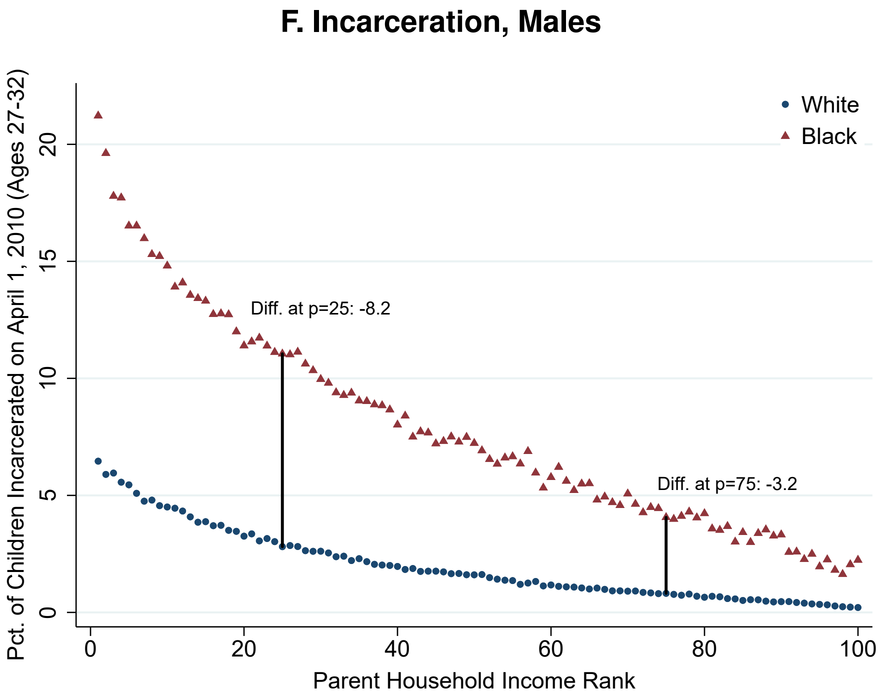


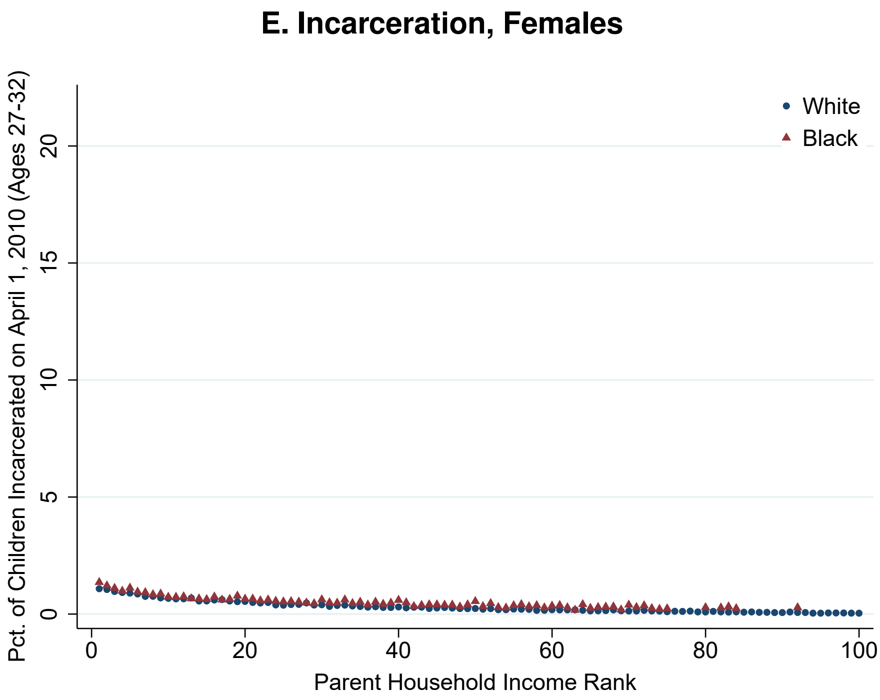

B. High School Completion Rates, Males
Pct. of Children with High School Degree (Age >= 19)
100
90
80
70
60
50
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=25: 8.3
Diff. at p=75: 4.2
White
Black


A. High School Completion Rates, Females
Pct. of Children with High School Degree (Age >= 19)
100
90
80
70
60
50
0
20
40
60
80
100
Parent Household Income Rank
Diff. at p=25: 3.5
Diff. at p=75: 1.5
White
Black


D. College Attendance Rates, Males
College Attendance Rate for Children (%)
Parent Household Income Rank
Diff. at p=25: 6.5
Diff. at p=75: 7.7
White
Black
100
80
60
40
20
0
20
40
60
80
100


C. College Attendance Rates, Females
College Attendance Rate for Children (%)
Parent Household Income Rank
Diff. at p=75: 3.6
Diff. at p=25: 2.8
White
Black
20
40
60
80
100
0
20
40
60
80
100


F. Incarceration, Males
Pct. of Children Incarcerated on April 1, 2010 (Ages 27-32)
Parent Household Income Rank
White
Black
Diff. at p=25: -8.2
Diff. at p=75: -3.2
0
5
10
15
20
0
20
40
60
80
100


E. Incarceration, Females
Pct. of Children Incarcerated on April 1, 2010 (Ages 27-32)
0
5
10
15
20
0
20
40
60
80
100
Parent Household Income Rank
White
Black

Experimental and Non-Experimental Approaches

- Observational decompositions provide valuable descriptive evidence, but suffer from the usual issues with non-experimental research designs
- Link between Oaxaca and treatment effect estimation under CIA is useful for thinking about potential issues
- Oaxaca decompositions can overstate or understate the extent of discrimination
  - Productivity-relevant characteristics that differ across groups may be excluded from X_i (= omitted variable bias)
  - Discrimination may affect the distribution of X_i (= bad control)
- Motivating by these issues, a parallel strand of literature uses experimental approaches to measure discrimination

Audit and Correspondence Studies

- Common experimental approach to studying discrimination: **audit** and **correspondence** studies
- In-person audit studies: send pairs of auditors matched on personal characteristics but different on some dimension of interest (e.g. race) to apply for a real job
- Resume correspondence studies: send fictitious resumes to real jobs with randomly assigned names that signify protected characteristics
  - Pioneered by Bertrand and Mullainathan (2004) to study racial discrimination
  - Baert (2018) counts 90 correspondence studies on hiring discrimination since 2005
- See Bertrand and Duflo (2017) for a survey of audit and correspondence studies in economics

Bertrand and Mullainathan (2004)

- BM sent four fake resumes to (almost) every employment ad posted in the Boston Globe or Chicago Tribune between summer 2001 and spring 2002 in sales, administrative support, clerical and customer services
  - Choose racially distinctive names based on empirical likelihood ratios among all children born in MA 1974-79
  - Resume "bank" built from characteristics of real resumes posted on job search web sites
  - Every job receives two resumes with white names and two resumes with black names
  - Main outcome: Employer callback

TABLE A1—FIRST NAMES USED IN EXPERIMENT

White female Name	L(W)/L(B)	Perception White	African-American female Name	L(B)/L(W)	Perception Black
Allison	∞	0.926	Aisha	209	0.97
Anne	∞	0.962	Ebony	∞	0.9
Carrie	∞	0.923	Keisha	116	0.93
Emily	∞	0.925	Kenya	∞	0.967
Jill	∞	0.889	Lakisha	∞	0.967
Laurie	∞	0.963	Latonya	∞	1
Kristen	∞	0.963	Latoya	∞	1
Meredith	∞	0.926	Tamika	284	1
Sarah	∞	0.852	Tanisha	∞	1
Fraction of all births:			Fraction of all births:		
3.8 percent			7.1 percent		

White male			African-American male		
Name	L(W)/L(B)	Perception White	Name	L(B)/L(W)	Perception Black
Brad	∞	1	Darnell	∞	0.967
Brendan	∞	0.667	Hakim		0.933
Geoffrey	∞	0.731	Jamal	257	0.967
Greg	∞	1	Jermaine	90.5	1
Brett	∞	0.923	Kareem	∞	0.967
Jay	∞	0.926	Leroy	44.5	0.933
Matthew	∞	0.888	Rasheed	∞	0.931
Neil	∞	0.654	Tremayne	∞	0.897
Todd	∞	0.926	Tyrone	62.5	0.900
Fraction of all births:			Fraction of all births:		
1.7 percent			3.1 percent		

TABLE 1—MEAN CALLBACK RATES BY RACIAL SOUNDINGNESS OF NAMES

	Percent callback for White names	Percent callback for African-American names	Ratio	Percent difference (p-value)
Sample:				
All sent resumes	9.65	6.45	1.50	3.20
	[2,435]	[2,435]		(0.0000)
Chicago	8.06	5.40	1.49	2.66
	[1,352]	[1,352]		(0.0057)
Boston	11.63	7.76	1.50	4.05
	[1,083]	[1,083]		(0.0023)
Females	9.89	6.63	1.49	3.26
	[1,860]	[1,886]		(0.0003)
Females in administrative jobs	10.46	6.55	1.60	3.91
	[1,358]	[1,359]		(0.0003)
Females in sales jobs	8.37	6.83	1.22	1.54
	[502]	[527]		(0.3523)
Males	8.87	5.83	1.52	3.04
	[575]	[549]		(0.0513)

TABLE 2—DISTRIBUTION OF CALLBACKS BY EMPLOYMENT AD

Equal Treatment:	No Callback	1W + 1B	2W + 2B
88.13 percent	83.37	3.48	1.28
[1,166]	[1,103]	[46]	[17]
Whites Favored (WF):	1W + 0B	2W + 0B	2W + 1B
8.39 percent	5.59	1.44	1.36
[111]	[74]	[19]	[18]
African-Americans Favored (BF):	1B + 0W	2B + 0W	2B + 1W
3.48 percent	2.49	0.45	0.53
[46]	[33]	[6]	[7]
Ho: $WF = BF$			
$p = 0.0000$			

TABLE 8—CALLBACK RATE AND MOTHER'S EDUCATION BY FIRST NAME

White female			African-American female		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Emily	7.9	96.6	Aisha	2.2	77.2
Anne	8.3	93.1	Keisha	3.8	68.8
Jill	8.4	92.3	Tamika	5.5	61.5
Allison	9.5	95.7	Lakisha	5.5	55.6
Laurie	9.7	93.4	Tanisha	5.8	64.0
Sarah	9.8	97.9	Latoya	8.4	55.5
Meredith	10.2	81.8	Kenya	8.7	70.2
Carrie	13.1	80.7	Latonya	9.1	31.3
Kristen	13.1	93.4	Ebony	9.6	65.6
Average		91.7	Average		61.0
Overall		83.9	Overall		70.2
Correlation	−0.318	(p = 0.404)	Correlation	−0.383	(p = 0.309)

White male			African-American male		
Name	Percent callback	Mother education	Name	Percent callback	Mother education
Todd	5.9	87.7	Rasheed	3.0	77.3
Neil	6.6	85.7	Tremayne	4.3	—
Geoffrey	6.8	96.0	Kareem	4.7	67.4
Brett	6.8	93.9	Darnell	4.8	66.1
Brendan	7.7	96.7	Tyrone	5.3	64.0
Greg	7.8	88.3	Hakim	5.5	73.7
Matthew	9.0	93.1	Jamal	6.6	73.9
Jay	13.4	85.4	Leroy	9.4	53.3
Brad	15.9	90.5	Jermaine	9.6	57.5
Average		91.7	Average		66.7
Overall		83.5	Overall		68.9
Correlation	−0.0251	(p = 0.949)	Correlation	−0.595	(p = 0.120)

Critiques of Audit Studies

- Audit and correspondence studies have been criticized for a variety of reasons (e.g. by Heckman and Siegelman, 1993 and Heckman, 1998)
  - Demand effects in in-person audit experiments
  - Behavior of average vs. marginal firms
  - Distinctively black names may signify attributes other than race (Gaddis, 2017)
  - Economic significance of callback outcome
  - Effects may be due to statistical rather than taste-based discrimination (do we care?)
- As noted by Guryan and Charles (2013), despite issues with interpretation, results from correspondence studies appear to demonstrate firms illegally use protected characteristics in the hiring process

Variation in Discrimination: Kline and Walters (forthcoming)

- Correspondence studies typically focus on market-level averages of discrimination
- Distribution of discrimination across employers is important for both research and policy
  - Economic models imply equilibrium impact of discrimination depends on prejudice of marginal employer rather than the average (Becker, 1957)
  - Enforcement of anti-discrimination law requires identifying individual offenders (e.g., EEOC charges)
- Kline and Walters (forthcoming) revisit correspondence evidence to study variation across employers
- Basic idea: Correspondence studies sending multiple applications per job provide a window into employer heterogeneity

Kline and Walters (forthcoming): Bernoulli Trials

- Starting point is a model for callbacks as independent Bernoulli trials
- Potential outcomes of application $i \in \{1, ..., N\}$ to job $j \in \{1, ..., J\}$ as a function of race $r \in \{b, w\}$:

$$Y_{ij}(r) \overset{iid}{\sim} Bernoulli(p_{jr})$$

- Key restriction is iid assumption: repeated trials at the same job are draws from a stable callback process
- A job is defined by its callback probabilities, p_{jb} and p_{jw}
- Discriminators have $p_{jb} \neq p_{jw}$

Kline and Walters (forthcoming): Hierarchical Model

- Independent trials implies callback counts C_{jw} and C_{jb} are binomial:

$$f(C_{jw}, C_{jb}|p_{jw}, p_{jb}) = \binom{N_w}{C_{jw}} p_{jw}^{C_{jw}} (1 - p_{jw})^{N_w - C_{jw}} \times \binom{N_b}{C_{jb}} p_{jb}^{C_{jb}} (1 - p_{jb})^{N_b - C_{jb}}$$

- Next, think about the joint distribution of p_{jw} and p_{jb} across jobs:

$$(p_{jw}, p_{jb}) \sim G(p_w, p_b)$$

- This is a **hierarchical model**
  - Binomial trials for each job
  - Heterogeneous success probabilities across jobs

Kline and Walters (forthcoming): Importance of $G(\cdot)$

- Distribution function $G(\cdot)$ describes heterogeneity in callback levels and discrimination. Share of jobs that discriminate is:

$$\bar{\pi} = \int_{p_w \neq p_b} dG(p_w, p_b)$$

- $G(\cdot)$ can also help us interpret evidence for individual jobs
- By Bayes' rule, share of non-discriminators among jobs with callback counts (C_{jw}, C_{jb}) is:

$$\Pr[p_{jw} \neq p_{jb} | C_{jw}, C_{jb}] = \frac{f(C_{jw}, C_{jb} | p_{jw} \neq p_{jb}) \Pr(p_{jw} \neq p_{jb})}{f(C_{jw}, C_{jb})}$$

$$= \frac{\int_{p_w \neq p_b} f(C_{jw}, C_{jb} | p_w, p_b) dG(p_w, p_b) \bar{\pi}}{f(C_{jw}, C_{jb})}$$

- If we knew $G(\cdot)$, we could calculate this probability
- **Empirical Bayes** approach: plug in an estimator $\hat{G}(\cdot)$ of the cross-job distribution to form posteriors for individual jobs

Kline and Walters (forthcoming): Identification of $G(\cdot)$

- It turns out that some features of $G(\cdot)$ are identified with only a few applications per job
- Share of jobs with callback counts (c_w, c_b):

$$f(c_w, c_b) = \binom{N_w}{c_w}\binom{N_b}{c_b} E\left[p_{jw}^{c_w}(1-p_{jw})^{N_w-c_w} p_{jb}^{c_b}(1-p_{jb})^{N_b-c_b}\right]$$

$$= \binom{N_w}{c_w}\binom{N_b}{c_b} \sum_{m=0}^{N_w-c_w} \sum_{n=0}^{N_b-c_b} (-1)^{m+n} \binom{N_w-c_w}{m}\binom{N_b-c_b}{n} E\left[p_{jw}^{c_w+m} p_{jb}^{c_b+n}\right]$$

- Unconditional callback probabilities are functions of moments of $G(\cdot)$
- Can solve for all moments $E[p_{jw}^m p_{jb}^n]$ for $0 \leq m \leq N_w$ and $0 \leq n \leq N_b$
- With two or more apps per race at each job, can identify measures of heterogeneity, e.g. $Var(p_{jb} - p_{jw})$

Table III.A: Treatment effect variation in BM (2004)

	p_b (1)	p_w (2)	$p_b - p_w$ (3)
Mean	0.063	0.094	-0.031
	(0.006)	(0.007)	(0.006)
Standard deviation	0.152	0.199	0.082
	(0.012)	(0.012)	(0.016)
Correlation with p_w or p_f	0.927	1.00	-0.717
	(0.051)	-	(0.119)

Table III.B: Treatment effect variation in NPRS (2015)

	p_b (1)	p_w (2)	$p_b - p_w$ (3)
Mean	0.153	0.177	-0.023
	(0.007)	(0.007)	(0.005)
Standard deviation	0.290	0.308	$\boxed{0.102}$
	(0.008)	(0.007)	$\boxed{(0.012)}$
Correlation with p_w or p_f	0.944	1.00	-0.336
	(0.017)	-	(0.066)
Skewness	3.76	3.65	$\boxed{-4.45}$
	(0.08)	(0.08)	$\boxed{(0.82)}$

Kline and Walters (forthcoming): Bounds

- Calculating discrimination probabilities requires more than a few moments of $G(\cdot)$
- But we can use the moments we have to calculate bounds
- Let $\mu(G)$ denote list of moments for distribution $G(\cdot)$, and f the list of observed callback probabilities
- Lower bound on the share of jobs that discriminate:

$$\bar{\pi} \geq \min_{G} \int_{p_w \neq p_b} dG(p_w, p_b) \text{ s.t. } f = B\mu(G)$$

- Discretize $G \implies$ this is a tractable linear programming problem
- We can use this approach to bound the overall share of discriminators, and the share of jobs with particular callback configurations that discriminate

Figure I: Lower bounds on posterior probabilities of discrimination, BM data

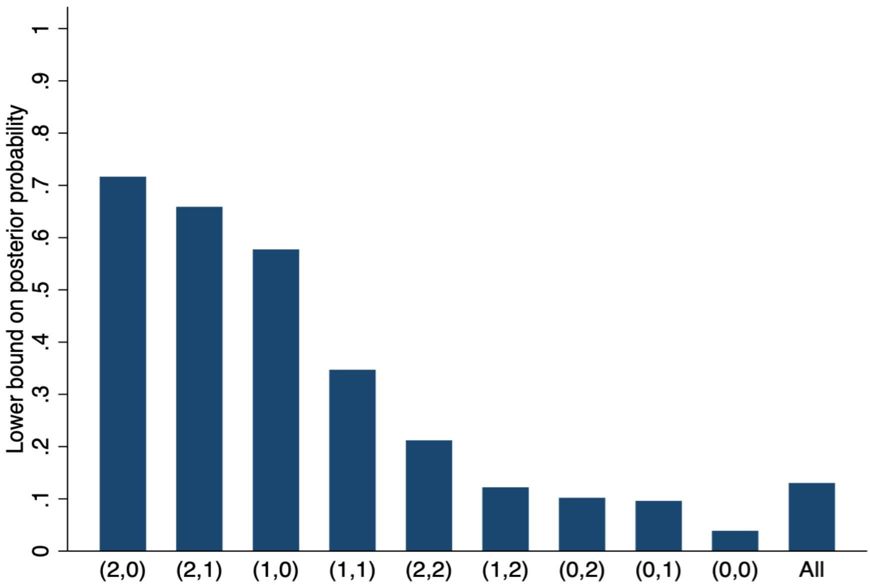
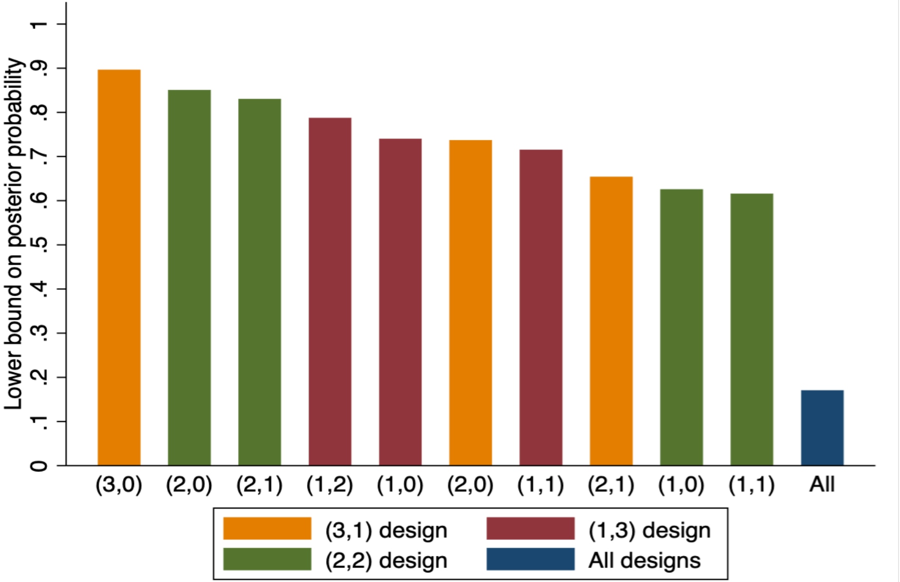


Figure II: Lower bounds on posterior probabilities of discrimination, Nunley et al. data



Kline and Walters (forthcoming): Decisions

- Results so far suggest it is possible to obtain informative posteriors for some individual jobs
- Can we use correspondence evidence to make decisions about which employers to investigate?
- Let $\delta(C_{jw}, C_{jb}) \in \{0, 1\}$ indicate decision to investigate as a function of callbacks
- Consider a simple **loss function:**

$$\mathcal{L}_j(\delta(C_{jw}, C_{jb})) = \gamma_1 \delta(C_{jw}, C_{jb}) 1\{p_{jw} = p_{jb}\} + \gamma_2 (1 - \delta(C_{jw}, C_{jb})) 1\{p_{jw} \neq p_{jb}\}$$

- γ_1 and γ_2 reflect costs of type I and type II errors
- Optimal decision rule minimizes **risk** (expected loss):

$$\delta^*(C_{jw}, C_{jb}) = \arg\min_{\delta(\cdot)} E\left[\mathcal{L}_j(\delta(C_{jw}, C_{jb}))\right]$$

Kline and Walters (forthcoming): Decisions

- Optimal decision is to investigate when posterior exceeds a cost-based threshold:

$$\delta^*(C_{jw}, C_{jb}) = 1\left\{\Pr[p_{wj} \neq p_{bj}|C_{jw}, C_{jb}] > \frac{\gamma_1}{\gamma_1+\gamma_2}\right\}.$$

- Unlike frequentist hypothesis testing, posterior threshold rule controls the **false discovery rate**, $FDR \equiv \Pr\left[p_{wj} = p_{bj}|\delta(C_{jw}, C_{jb}) = 1\right]$

- Close link to literature on multiple testing (Benjamini and Hochberg, 1995; Storey, 2002; Efron, 2012)

- With knowledge of $G(\cdot)$, can trace out tradeoff between type I and II errors (**detection/error tradeoff** curve)

Kline and Walters (forthcoming): Parametric Model

- Study detection/error tradeoffs with a parametric model for $G(\cdot)$
- Mixed logit model for callback to application i at job j:

$$\Pr\left(Y_{ij}=1|\alpha_j,\beta_j,R_{ij},X_{ij}\right)=\frac{\exp\left(\alpha_j-\beta_j 1\{R_{ij}=b\}+X_{ij}'\psi\right)}{1+\exp\left(\alpha_j-\beta_j 1\{R_{ij}=b\}+X_{ij}'\psi\right)}.$$

- R_{ij} indicates race, $X_{j\ell}$ includes other randomly-assigned characteristics (GPA, experience, etc.)
- Two-type mixing:

$$\alpha_j \sim N\left(\alpha_0,\sigma_\alpha^2\right),$$

$$\beta_j=\begin{cases}\beta_0, & \text{with prob. } \frac{\exp(\tau_0+\tau_\alpha\alpha_j)}{1+\exp(\tau_0+\tau_\alpha\alpha_j)}\\ 0, & \text{with prob. } \frac{1}{1+\exp(\tau_0+\tau_\alpha\alpha_j)}.\end{cases}$$

- Kline and Walters (forthcoming) also consider decisions with continuous loss and partial identification of $G(\cdot)$ (**minimax** analysis)

Table V: Mixed logit parameter estimates, NPRS data

		Constant	Types	
			No selection	Selection
		(1)	(2)	(3)
Distribution of logit(p_w):	α_0	-4.71	-4.93	-4.93
		(0.22)	(0.24)	(0.28)
	σ_α	4.74	4.99	4.98
		(0.22)	(0.25)	(0.29)
Discrimination intensity:	β_0	0.456	$\boxed{4.05}$	4.05
		(0.108)	$\boxed{(1.56)}$	(1.58)
Discrimination logit:	τ_0	-	-1.59	-1.56
			(0.42)	(1.10)
	τ_α	-	-	-0.005
				(0.180)
Fraction with $p_w \neq p_b$:		1.00	$\boxed{0.168}$	0.170
Log-likelihood		-2,792.1	-2,788.2	-2,788.2
Parameters		15	16	17
Sample size		2,305	2,305	2,305

Figure V: Detection/error tradeoffs, NPRS data

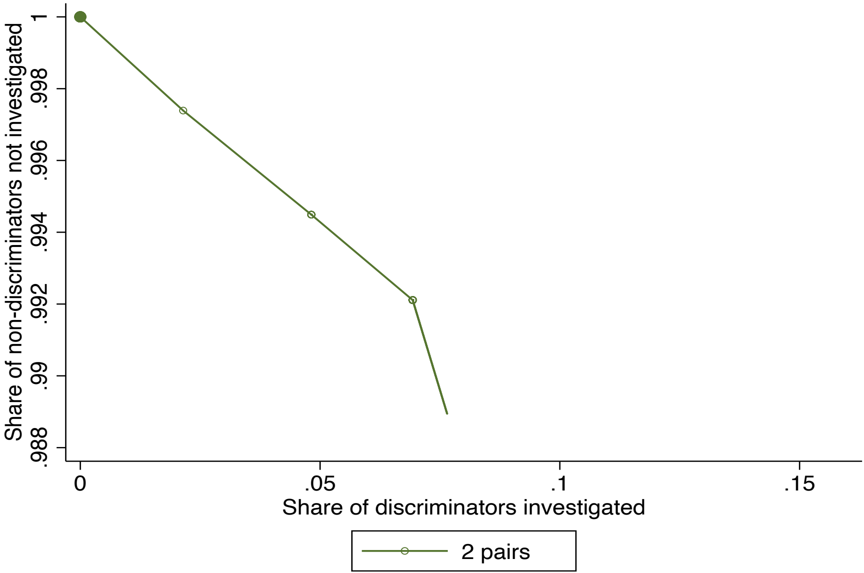


Figure V: Detection/error tradeoffs, NPRS data

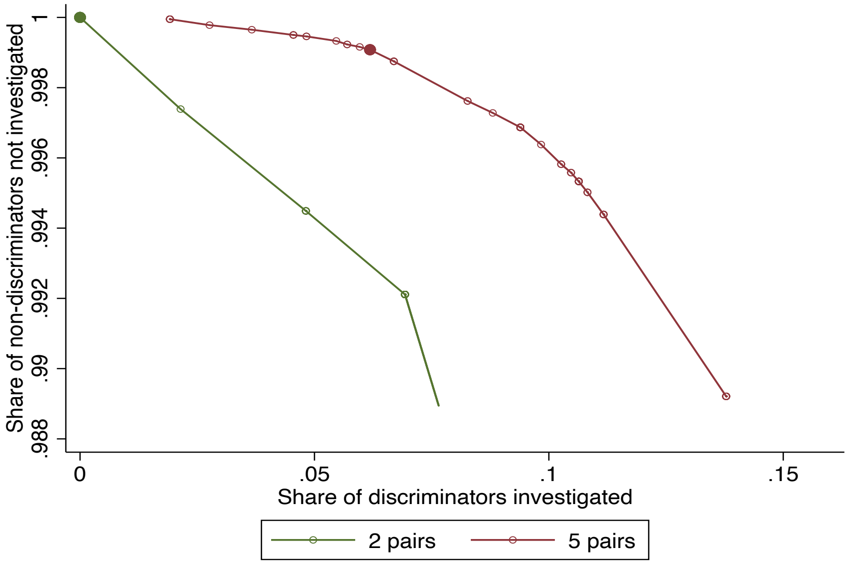


Figure V: Detection/error tradeoffs, NPRS data

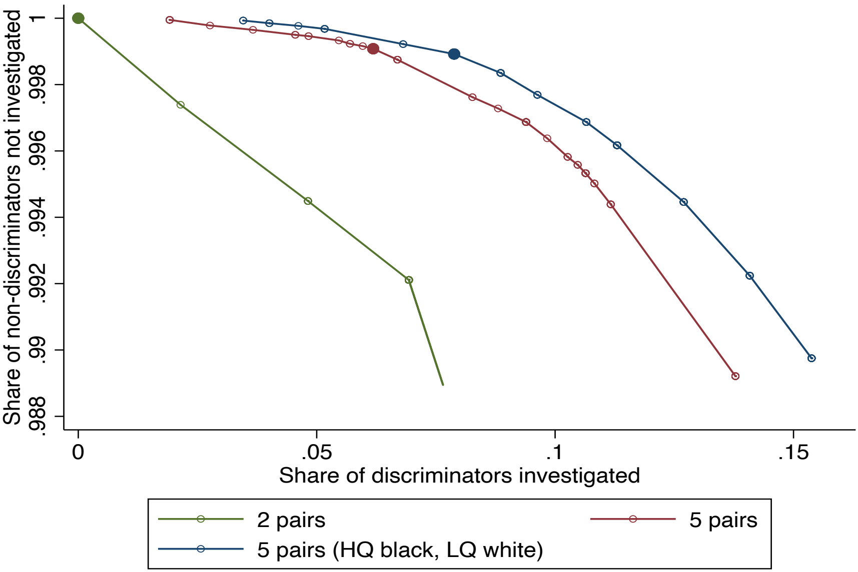
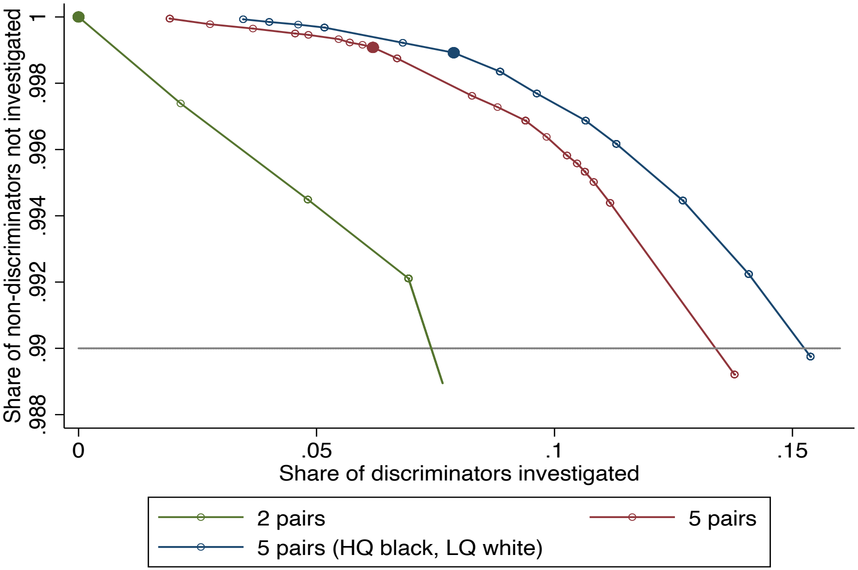


Figure V: Detection/error tradeoffs, NPRS data



Disparate Impacts of Probation: Rose (forthcoming)

- Large racial disparities in criminal justice outcomes motivate studies of discrimination in the criminal justice system
- Concerns that formally neutral policies may have disproportionate effects on protected groups
  - Disparate treatment vs. disparate impact
  - Algorithmic fairness and bias (Kleinberg et al., 2017; Yang and Dobbie, 2020)
- Rose (forthcoming QJE) studies the impacts of probation rules on racial gaps in incarceration

Rose (forthcoming): Background

- Most convicted offenders in the US serve sentences on probation ("community supervision") rather than in prison
- While on probation, violations of technical rules (e.g. failure to pay fees or fines) may result in incarceration
- Probation revocations account for 25% of prison admissions, and are more common among black probationers
- Technical rules are meant to serve two purposes:
  - Support reintegration/rehabilitation of offenders
  - Identify those who are likely to offend again
- Key question: Do probation violations accurately target high-risk offenders?

Rose (forthcoming): Probation reform

- Rose (forthcoming) studies a 2011 North Carolina reform that reduced punishments for technical probation violations
- Prior to reform, revocations for nonpayment of fees/fines or failing drug/alcohol tests were common
- Afterwards, revocations were reserved for new crimes or absconding (fleeing supervision)
- Using administrative data from NC, Rose (forthcoming) employs a difference-in-differences design to study the impacts of this reform
    - Treatment group: Offenders on supervised probation in NC
    - Control group: unsupervised probationers (less serious offenses/not monitored for violations)

Difference-in-Differences

- Consider individuals in two groups $g \in \{A, B\}$ observed in two time periods $t \in \{pre, post\}$
- Treatment switches on for group A in the post period: $D_{igt} = 1\{g = A, t = post\}$
- Let $Y_{igt}(d)$ denote potential outcome for individual i in group g in period t with treatment status $d \in \{0, 1\}$
- We observe $E[Y_{i,A,post}(1)]$, and $E[Y_{igt}(0)]$ for $(g, t) \in \{(A, 0), (B, 0), (B, 1)\}$
- Objective: calculate the treatment effect for group A in the post period, $E[Y_{i,A,post}(1) - Y_{i,A,post}(0)]$
- Requires imputing the unobserved counterfactual for the treated group in the post period, $E[Y_{i,A,post}(0)]$

Difference-in-Differences

- The core of a diff-in-diff design is an additive model for non-treated potential outcomes:

$$E[Y_{igt}(0)] = \alpha_g + \gamma_t$$

- α_g is a time-invariant group effect
- γ_t is a group-invariant time effect
- Groups can be different, and time periods can be different. Key is no time-varying group-specific confounders
- Additive model implies **parallel trends** in non-treated potential outcomes across groups:

$$E[Y_{i,A,post}(0)] - E[Y_{i,A,pre}(0)] = E[Y_{i,B,post}(0)] - E[Y_{i,B,pre}(0)]$$

$$= \gamma_1 - \gamma_0.$$

Difference-in-Differences

- With an additive model, the observed change in outcomes for group B captures the counterfactual change in outcomes for group A, so:

$$E[Y_{i,A,post}(0)] = E[Y_{i,A,pre}(0)] + (E[Y_{i,B,post}(0)] - E[Y_{i,B,pre}(0)])$$

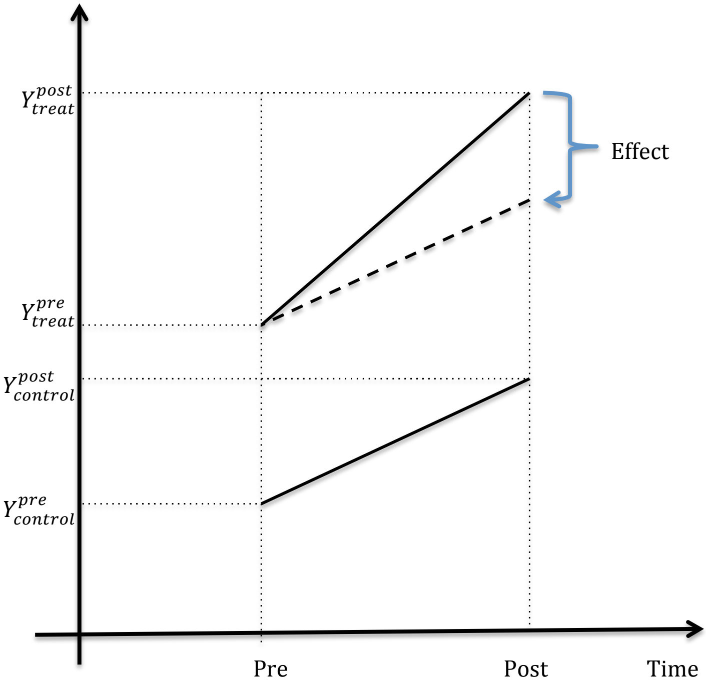
- The treatment effect for group A in the post period is then:

$$E[Y_{i,A,post}(1) - Y_{i,A,post}(0)] = (E[Y_{i,A,post}(1)] - E[Y_{i,A,pre}(0)])$$
$$- (E[Y_{i,B,post}(0)] - E[Y_{i,B,pre}(0)])$$

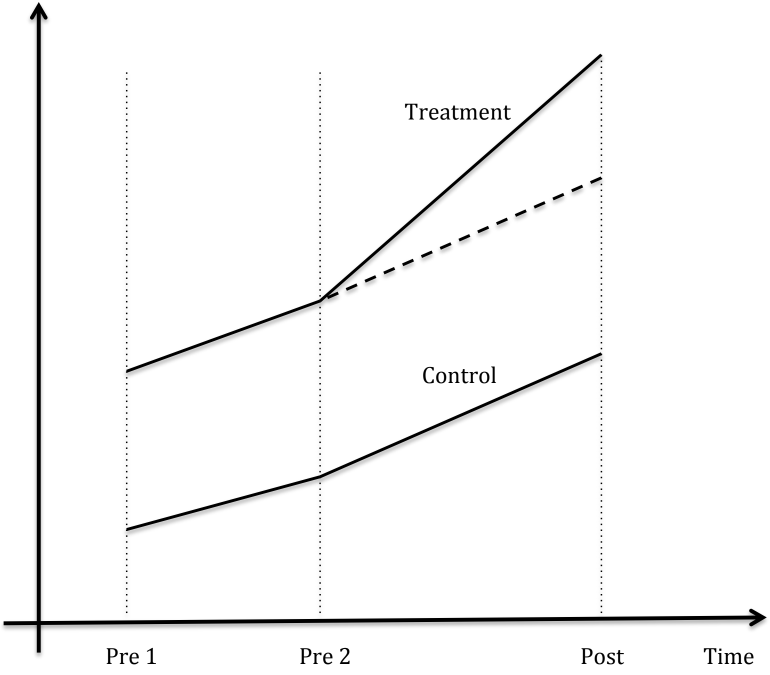
- Diff-in-diff looks at the change in outcomes for the treated group, subtracting off the change for the control group to eliminate time effects
- Implement with linear regression:

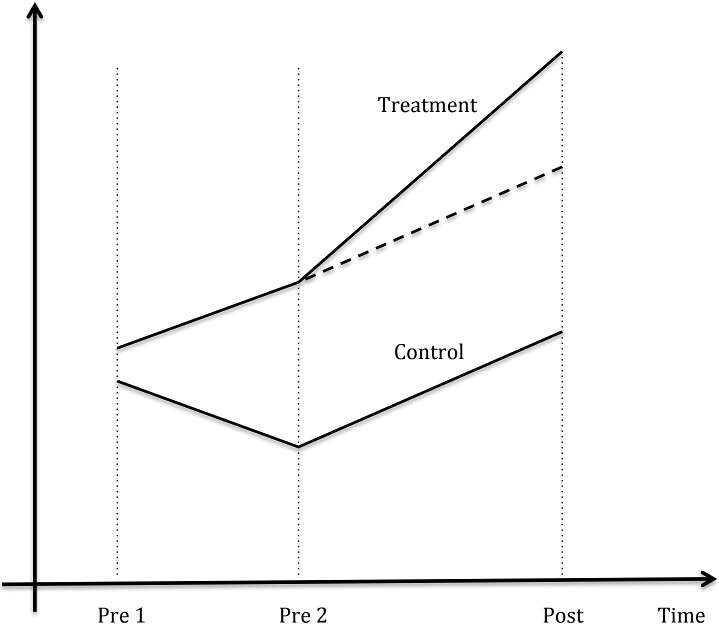
$$Y_{igt} = \alpha_g + \gamma_t + \beta D_{gt} + \epsilon_{igt}$$

- Coefficient β captures difference in changes over time for treatment vs. control



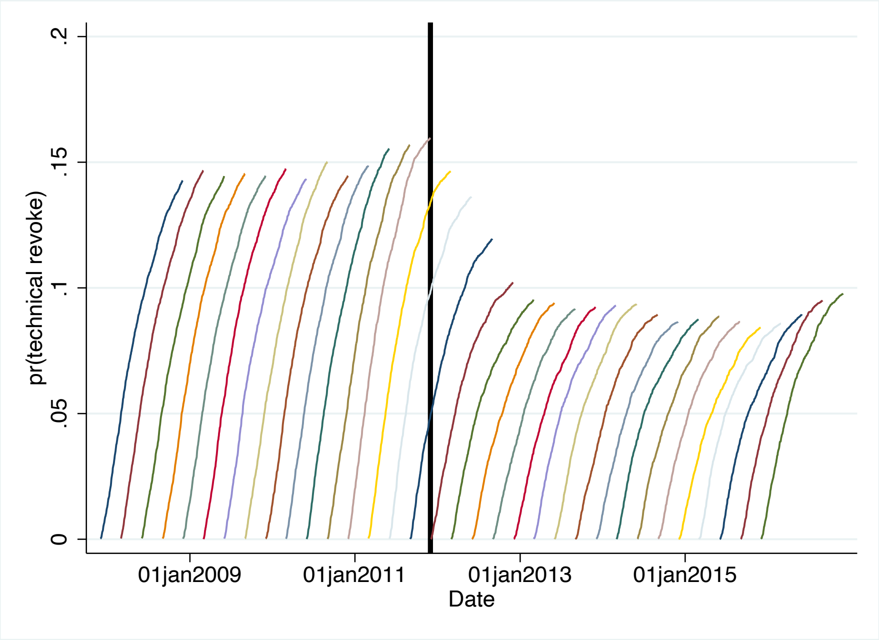
Y_{treat}^{post}
Effect
Y_{treat}^{pre}
$Y_{control}^{post}$
$Y_{control}^{pre}$
Pre
Post
Time


Treatment
Control
Pre 1
Pre 2
Post
Time



Treatment
Control
Pre 1
Pre 2
Post
Time

A. Technical revocation



B. Arrests

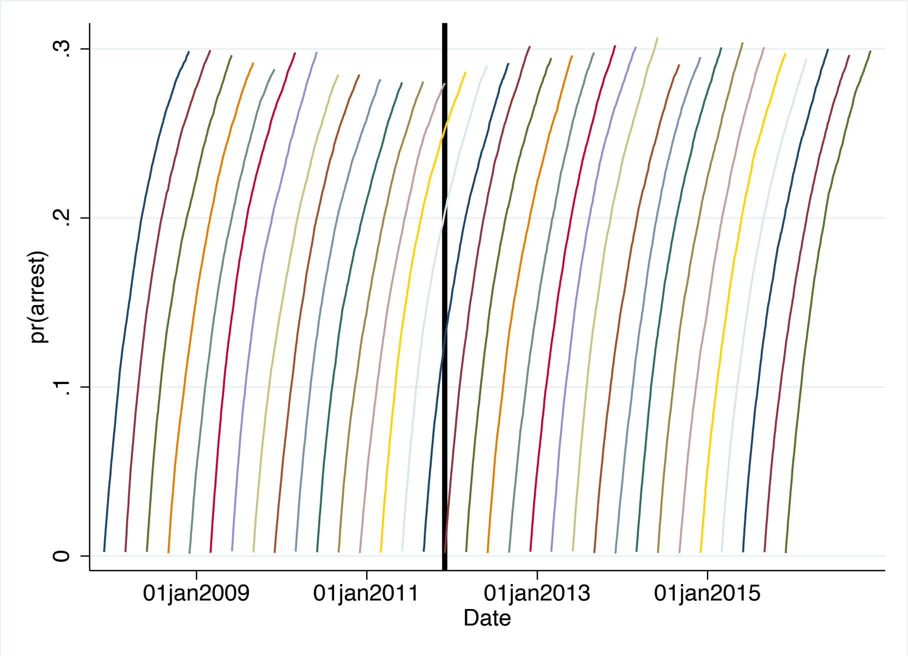
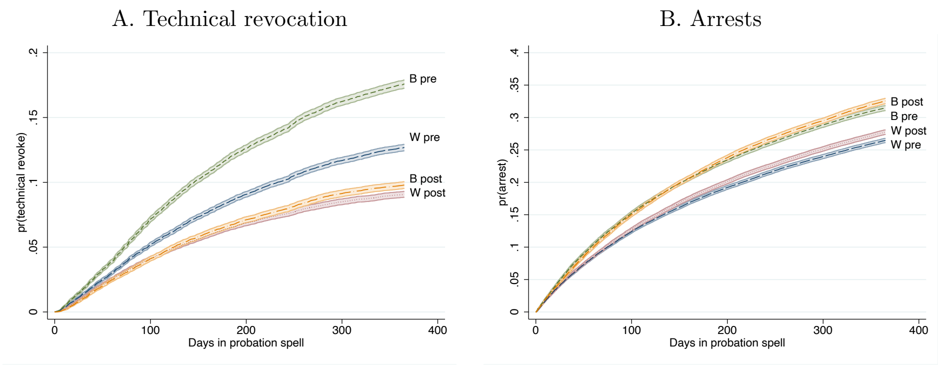

pr(arrest)
0
.1
.2
.3
01jan2009
01jan2011
01jan2013
01jan2015
Date

FIGURE III
EFFECTS OF REFORM BY RACE



Notes. This figure plots effects of the 2011 JRA reform on technical revocation and arrests separately by race. It includes all supervised probationers starting their spells either 1-3 years before (pre) or 0-2 years after the reform (post). "B" refers to black probationers, while "W" refers to non-black. The y-axis measures the share of each group experiencing the relevant outcome over the first year of their probation spell. Technical revocation is an indicator for having probation revoked for rule violations with no intervening criminal arrest. Arrest is an indicator for a criminal arrest before revocation for any rule violations. Shaded areas reflect 95% confidence intervals formed using standard errors clustered at the individual level.

TABLE III

Difference-in-Differences Estimates of Reform Impacts

A. All offenders

	Technical revoke		Arrest	
	(1)	(2)	(3)	(4)
Post-reform	-0.00172***	-0.00205***	-0.00793***	-0.00705***
	(0.000273)	(0.000288)	(0.00167)	(0.00159)
Treated	0.143***	0.133***	0.0316***	-0.0155***
	(0.00103)	(0.00102)	(0.00166)	(0.00164)
Post-x-treat	-0.0532***	-0.0530***	0.0196***	0.0194***
	(0.00135)	(0.00135)	(0.00242)	(0.00233)
N	546006	546006	546006	546006

B. Non-black offenders

Post-reform	-0.000522	-0.000875**	-0.00693***	-0.00666***
	(0.000317)	(0.000334)	(0.00199)	(0.00190)
Treated	0.122***	0.112***	0.0450***	-0.000334
	(0.00130)	(0.00126)	(0.00209)	(0.00207)
Post-x-treat	-0.0356***	-0.0360***	0.0198***	0.0179***
	(0.00173)	(0.00172)	(0.00304)	(0.00295)
N	328784	328784	328784	328784

C. Black offenders

Post-reform	-0.00387***	-0.00412***	-0.0118***	-0.0112***
	(0.000509)	(0.000534)	(0.00295)	(0.00281)
Treated	0.167***	0.160***	-0.00496	-0.0464***
	(0.00167)	(0.00167)	(0.00274)	(0.00268)
Post-x-treat	-0.0741***	-0.0736***	0.0228***	0.0233***
	(0.00215)	(0.00214)	(0.00399)	(0.00383)
N	217222	217222	217222	217222

Rose (forthcoming): Interpretation

- Rose finds that eliminating technical violations led to large reductions in probation revocation and modest increases in re-offending
- Effects on revocation are much larger for black probationers than white probationers, while effects on re-offending are comparable
- Reform eliminated large racial gap in revocations with no impact on racial gap in re-offending
- This suggests technical revocations target re-offenders more accurately for white probationers

Rose (forthcoming): Targeting

- To formally analyze targeting accuracy, ignore time dimension and consider LATE setup:
  - Z_i: indicator equal to one if i is subject to technical rules
  - $R_i(1)$, $R_i(0)$: i's potential revocations as a function of Z_i
  - $Y_i(1)$, $Y_i(0)$: i's potential re-offending as a function of R_i
- Probability of a technical revocation among those who would not otherwise be revoked:

$$\Pr(R_i(1) = 1 | R_i(0) = 0)$$

- This is the share of compliers ($R_i(1) > R_i(0)$) among the population of compliers and never takers ($R_i(0) = 0$)

Rose (forthcoming): Targeting

- By the law of total probability, we can write

$$\Pr(R_i(1) = 1 | R_i(0) = 0) = \mu_0\pi_0 + \mu_1\pi_1$$

- $\mu_k = \Pr(Y_i(0) = k | R_i(0) = 0)$ describes the distribution of re-offending risk among compliers and never-takers
- $\pi_0 = \Pr(R_i(1) = 1 | R_i(0) = 0, Y_i(0) = 0)$: False positive rate (Type I error)
- $\pi_1 = \Pr(R_i(1) = 1 | R_i(0) = 0, Y_i(0) = 1)$: True positive rate (one minus Type II error)
- Note that under the LATE model assumptions, all of these terms are identified

Rose (forthcoming): Targeting

- Oaxaca decomposition of racial difference in revocation rates:

$$\Pr(R_i(1) = 1 | R_i(0) = 0, B_i = 1) - \Pr(R_i(1) = 1 | R_i(0) = 0, W_i = 1)$$

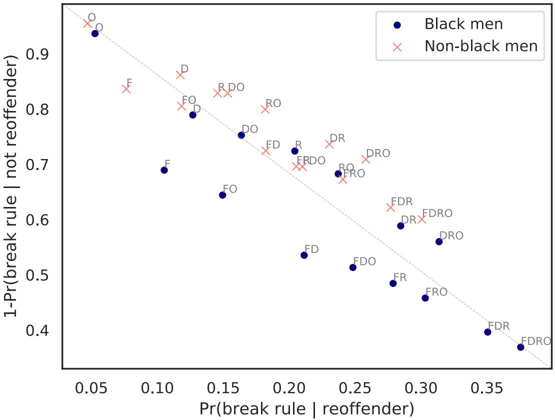
$$= \underbrace{\sum_{k=0}^{1} \mu_{k,W}(\pi_{k,B} - \pi_{k,W})}_{\textit{Targeting}} + \underbrace{\sum_{k=0}^{1} (\mu_{k,B} - \mu_{k,W})\pi_{k,B}}_{\textit{Risk}}$$

- First term is due to differences in targeting accuracy (true/false positive rates)
- Second term is due to differences in re-offending risk

TABLE IV
Decomposition of Racial Gaps in Revocations

	Overall rates		Decomposition	
	White	Black	Gap	Share of gap explained
Probability of technical revoke:				
$Pr(R(1) = 1)$	0.039	0.082	0.043	100.0%
Distribution of risk:				
$Pr(Y(0) = 1)$	0.313	0.376	0.063	9.8%
$Pr(Y(0) = 0)$	0.687	0.624	-0.063	-13.3%
True / false positive rates:				
$Pr(R(1)=1 \vert Y(0) = 1)$	0.070	0.068	-0.002	-1.5%
$Pr(\mathrm{R}(1)=1 \vert Y(0) = 0)$	0.025	0.091	0.066	104.9%

FIGURE V
EFFICIENCY AND EQUITY OF TECHNICAL VIOLATION RULE TYPES



Notes. This figure plots estimates of the share of potential reoffenders over a three year period who break technical rules before they reoffend (x-axis) against the share of non-reoffenders who do not break technical rules. Estimates come from simulating the model estimated in Section VD using a different set of rules. Each point is labeled with a combination of "F" for fees / fines violations, "D" for drug / alcohol violations, "R" for reporting violations, and "O" for all other, reflecting the sets of rules enforced in the simulation. The points labeled "FDRO" therefore reflect the set of rules punishable with incarceration before the 2011 reform, and "R" reflect the set punishable afterwards. The dotted gray line starts at $(1, 0)$ and has a slope of -1. This line reflects what would be achieved by randomly revoking a fraction of probationers at the start of their spells, which naturally would catch equal shares of reoffenders and non-reoffenders.

References

- Aigner, D., and Cain, G. (1977). "Statistical theories of discrimination in labor markets." *Industrial and Labor Relations Review* 30(2).
- Altonji, J., and Blank, R. (1999). "Race and gender in the labor market." Handbook of Labor Economics, Volume 3.
- Arrow, K. (1973). "The theory of discrimination." In: Discrimination in Labor Markets, Orley Ashenfelter and Albert Rees, eds.
- Baert, S. (2018). "Hiring discrimination: an overview of (almost) all correspondence experiments since 2005. In: Audit Studies: Behind the Scenes with Theory, Method, and Nuance. Springer.
- Becker, G. (1957). The Economics of Discrimination. University of Chicago Press.
- Benjamini, Y., and Hockberg, Y. (1995). "Controlling the false discovery rate: a practical annd powerful approach to multiple testing." *Journal of the Royal Statistical Society* 57.
- Bertrand, M., and Duflo, E. (2017). "Field experiments in discrimination." Handbook of Field Experiments, Volume 1.
- Bertrand, M., and Mullainathan, S. (2004). "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination." *American Economic Review* 94(4).

References

- Blinder, A. (1973). "Wage discrimination: reduced form and structural estimates." *Journal of Human Resources* 8(4).
- Chetty, R., Hendren, N., Jones, M., and Porter, S. (2020). "Race and economic opportunity in the United States: an intergenerational perspective." *Quarterly Journal of Economics* 135(2).
- Efron, B. (2012). Large-scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction. Cambridge University Press.
- Gaddis, S. (2017). "How black are Lakisha and Jamal? Racial perceptions from names used in correspondence audit studies." *Sociological Science* 4.
- Guryan, J., and Charles, K. (2013). "Taste-based or statistical discrimination: the economics of discrimination returns to its roots." *The Economic Journal* 123(572).
- Heckman, J., and Siegelman, P. (1993). "The Urban Institute audit studies: their methods and findings." In: Clear and Convincing Evidence: Measurement of Discrimination in America, M. Fix and R. Struyk, eds.
- Heckman, J. (1998). "Detecting discrimination." *Journal of Economic Perspectives* 12(2).
- Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., and Mullainathan, S. (2017). "Human decisions and machine predictions" *Quarterly Journal of Economics* 133(1).

References

- Kline, P. (2011). "Oaxaca-Blinder as a reweighting estimator." *American Economic Review Papers and Proceedings* 101(3).
- Kline, P., and Walters, C. (forthcoming). "Reasonable doubt: experimental detection of job-level employment discrimination." *Econometrica*.
- Lang, K., and Lehmann, J. (2012). "Racial discrimination in the labor market: theory and empirics." *Journal of Economic Literature* 50(4).
- Oaxaca, R. (1973). "Male-female wage differentials in urban labor markets." *International Economic Review* 14(3).
- Rose, E. (forthcoming). "Who gets a second chance? Effectiveness and equity in supervision of criminal offenders." *Quarterly Journal of Economics*.
- Storey, J. D. (2003). "A direct approach to false discovery rates," Journal of the Royal Statistical Society: Series B (Statistical Methodology) 64.
- Yang, C., and Dobbie, W. (2020). "Equal protection under algorithms: a new statistical and legal framework." *Michigan Law Review* 119(2).

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture #1: The Minimum Wage

Patrick Kline

UC Berkeley

Minimum wages: Background

Long presumption that minimum wages reduce employment (e.g., Stigler, 1946)

Evidence from time series / state panel regressions on aggregates typically find small but significant disemployment effects (Brown, 1999)

Some limitations of older literature

1. Exogeneity of min wage changes (who is the control group?)
2. Aggregates mask distributional impacts
3. Selective reporting bias

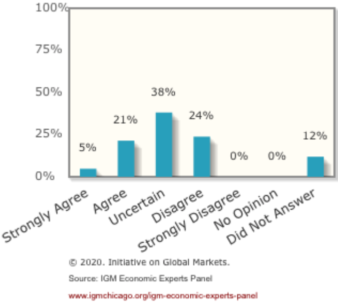
Important advances in methods and data in recent years have changed views on costs/benefits of min wage

Elite views circa Sept 2015

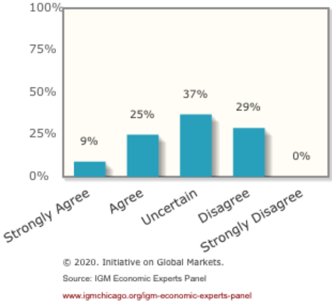
Question A:

If the federal minimum wage is raised gradually to $15-per-hour by 2020, the employment rate for low-wage US workers will be substantially lower than it would be under the status quo.

Responses



Responses weighted by each expert's confidence



Card and Krueger (1995): Do all t-stats = 2?

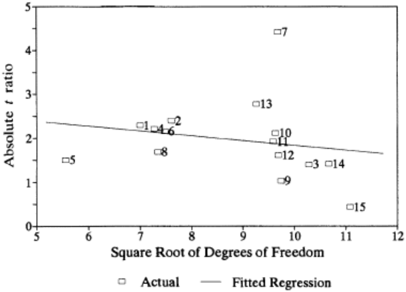


FIGURE 1. RELATION OF ESTIMATED *t* RATIOS TO SAMPLE SIZE

Note: The numbers in the figure refer to the following studies: 1) Hyman Kaitz, 1970; 2) Jacob Mincer, 1976; 3) Edward Gramlich, 1976; 4) Finis Welch, 1976; 5) James Ragan, 1977; 6) Michael Wachter and Choongsoo Kim, 1979; 7) George Iden, 1980; 8) Ragan, 1981; 9) John M. Abowd and Mark R. Killingsworth, 1981; 10) Charles Betsey and Bruce Dunson, 1981; 11) Brown et al., 1983; 12) Daniel Hamermesh, 1981; 13) Solon, 1985; 14) Wellington, 1991; 15) Klerman, 1992.

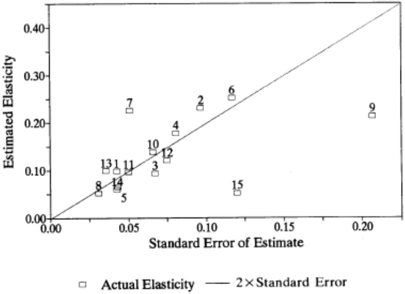


FIGURE 2. RELATION OF ESTIMATED EMPLOYMENT ELASTICITY TO STANDARD ERROR

Andrews and Kasy (2019): spikes at $t = 2$ a decade later

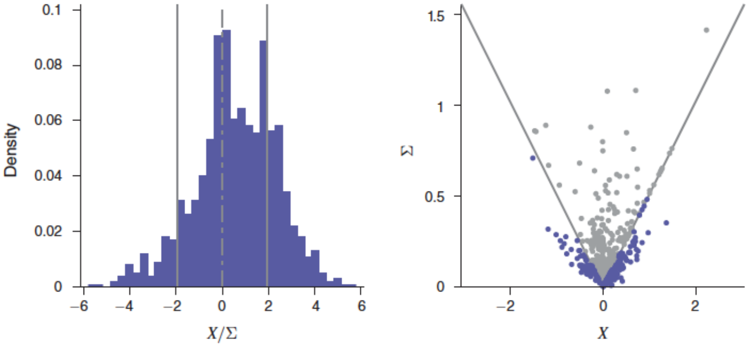


FIGURE 9. WOLFSON AND BELMAN (2015) DATA

Notes: The left panel shows a binned density plot for the z-statistics X/Σ in the Wolfson and Belman (2015) data. The solid gray lines mark $|X|/\Sigma = 1.96$, while the dash-dotted gray line marks $X/\Sigma = 0$. The right panel plots the estimate X against its standard error Σ. The gray lines mark $|X|/\Sigma = 1.96$.

Selection correct the t-stats

A selection model for reporting results

$$\Pr\left(\text{report} \mid Z\right) \propto \begin{cases} \beta_{p,1} & \text{if } Z < -1.96 \\ \beta_{p,2} & \text{if } -1.96 \leq Z < 0 \\ \beta_{p,3} & \text{if } 0 < Z \leq 1.96 \\ 1 & \text{if } Z < 1.96 \end{cases}$$

Selection bias when β_p's $<$1.

Latent population model of Z stats

$$\Theta \sim \bar{\theta} + t\left(\nu\right) \cdot \tau$$

where $\bar{\theta}$ is unselected mean, τ is scale parameter and ν is degrees of freedom for Student's t-distribution (low ν means fat tails)

Estimate by maximum likelihood treating studies as independent

Maximum likelihood estimates

TABLE 3—SELECTION ESTIMATES FOR WOLFSON AND BELMAN (2015)

$\bar{\theta}$	$\tilde{\tau}$	$\tilde{\nu}$	$\beta_{p,1}$	$\beta_{p,2}$	$\beta_{p,3}$
0.018	0.019	1.303	0.697	0.270	0.323
(0.009)	(0.011)	(0.279)	(0.350)	(0.111)	(0.094)

Notes: Meta-study estimates from minimum wage data, with standard errors clustered by study in parentheses. Publication probabilities β_p measured relative to omitted category of estimates positive and significant at 5 percent level.

- Severe selection: $\sim 30\%$ chance of reporting an insignificant result!
- Mean employment-MW elasticity borderline significant (the irony!)
  - Incredibly fat tails ($\nu < 2 \Rightarrow$variance doesn't exist!)
  - No accounting for study quality

Card and Krueger (1994): a trip down memory lane

Evaluate effects of April 1992 increase in NJ min wage from \$4.25 to \$5.05

Surveyed 410 fast-food restaurants in NJ and PA before and after change

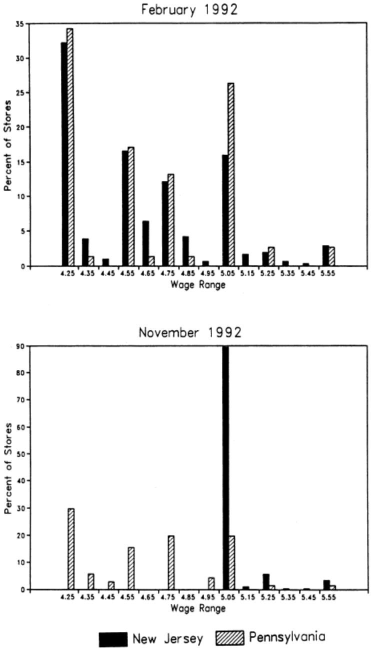
Two designs:

1. Diff in diff: compare NJ to PA
2. Exposure (gap) design: compare initially low wage to high wage establishments

Key findings:

- No (dis-)employment effect (possibly positive)
- Some evidence of cost pass-through to consumers

First stage looks good!



Two designs

- Diff in Diff

$$\Delta E_i = a + X_i' b + cNJ_i + \varepsilon_i$$

where X_i is baseline store characteristics

- Exposure design

$$\Delta E_i = \tilde{a} + \tilde{X}_i' \tilde{b} + \tilde{c}GAP_i + \tilde{\varepsilon}_i$$

where $\tilde{X}_i$ may include NJ_i and

$$GAP_i = NJ_i \cdot \max\left\{\frac{5.05 - W_{1i}}{W_{1i}}, 0\right\}$$

Zero or positive?

TABLE 4—REDUCED-FORM MODELS FOR CHANGE IN EMPLOYMENT

Independent variable	Model (i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	2.33 (1.19)	2.30 (1.20)	—	—	—
2. Initial wage gap[a]	—	—	15.65 (6.08)	14.92 (6.21)	11.91 (7.39)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	8.79	8.78	8.76	8.76	8.75
6. Probability value for controls[d]	—	0.34	—	0.44	0.40

Notes: Standard errors are given in parentheses. The sample consists of 357 stores with available data on employment and starting wages in waves 1 and 2. The dependent variable in all models is change in FTE employment. The mean and standard deviation of the dependent variable are −0.237 and 8.825, respectively. All models include an unrestricted constant (not reported).

[a]Proportional increase in starting wage necessary to raise starting wage to new minimum rate. For stores in Pennsylvania the wage gap is 0.

[b]Three dummy variables for chain type and whether or not the store is company-owned are included.

[c]Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

[d]Probability value of joint F test for exclusion of all control variables.

Do consumers pay more?

TABLE 7—REDUCED-FORM MODELS FOR CHANGE IN THE PRICE OF A FULL MEAL

Independent variable	Dependent variable: change in the log price of a full meal (i)	(ii)	(iii)	(iv)	(v)
1. New Jersey dummy	0.033 (0.014)	0.037 (0.014)	—	—	—
2. Initial wage gap[a]	—	—	0.077 (0.075)	0.146 (0.074)	0.063 (0.089)
3. Controls for chain and ownership[b]	no	yes	no	yes	yes
4. Controls for region[c]	no	no	no	no	yes
5. Standard error of regression	0.101	0.097	0.102	0.098	0.097

Notes: Standard errors are given in parentheses. Entries are estimated regression coefficients for models fit to the change in the log price of a full meal (entrée, medium soda, small fries). The sample contains 315 stores with valid data on prices, wages, and employment for waves 1 and 2. The mean and standard deviation of the dependent variable are 0.0173 and 0.1017, respectively.

[a]Proportional increase in starting wage necessary to raise the wage to the new minimum-wage rate. For stores in Pennsylvania the wage gap is 0.

[b]Three dummy variables for chain type and whether or not the store is company-owned are included.

[c]Dummy variables for two regions of New Jersey and two regions of eastern Pennsylvania are included.

Imprecise positive effects on store openings

TABLE 8—ESTIMATED EFFECT OF MINIMUM WAGES ON NUMBERS OF McDONALD'S RESTAURANTS, 1986–1991

	Dependent variable: proportional increase in number of stores				Dependent variable: (number of newly opened stores) ÷ (number in 1986)			
Independent variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
Minimum-Wage Variable:								
1. Fraction of retail workers in affected wage range 1986[a]	0.33 (0.20)	—	0.13 (0.19)	—	0.37 (0.22)	—	0.16 (0.21)	—
2. (State minimum wage in 1991) ÷ (average retail wage in 1986)[b]	—	0.38 (0.22)	—	0.47 (0.22)	—	0.47 (0.23)	—	0.56 (0.24)
Other Control Variables:								
3. Proportional growth in population, 1986–1991	—	—	0.88 (0.23)	1.03 (0.23)	—	—	0.86 (0.25)	1.04 (0.25)
4. Change in unemployment rates, 1986–1991	—	—	−1.78 (0.62)	−1.40 (0.61)	—	—	−1.85 (0.68)	−1.40 (0.65)
5. Standard error of regression	0.083	0.083	0.071	0.068	0.088	0.088	0.077	0.073

Notes: Standard errors are shown in parentheses. The sample contains 51 state-level observations (including the District of Columbia) on the number of McDonald's restaurants open in 1986 and 1991. The dependent variable in columns (i)–(iv) is the proportional increase in the number of restaurants open. The mean and standard deviation are 0.246 and 0.085, respectively. The dependent variable in columns (v)–(viii) is the ratio of the number of new stores opened between 1986 and 1991 to the number open in 1986. The mean and standard deviation are 0.293 and 0.091, respectively. All regressions are weighted by the state population in 1986.

[a]Fraction of all workers in retail trade in the state in 1986 earning an hourly wage between $3.35 per hour and the "effective" state minimum wage in 1990 (i.e., the maximum of the federal minimum wage in 1990 ($3.80) and the state minimum wage as of April 1, 1990).

[b]Maximum of state and federal minimum wage as of April 1, 1990, divided by the average hourly wage of workers in retail trade in the state in 1986.

The power of zero

A carefully thought out and transparent attempt to evaluate a min wage change w/ microdata

- Results a bit under-powered to detect clear positive but precise enough to reject big negative
- Inferential issue: no clustering
  - Debatable if we want to cluster (Abadie et al, 2020)
    - Are we conducting inference on the effect in this state or some hypothetical new state drawn from a super-population?
    - Does every DiD paper need to be a meta-analysis?
  - Either way dependence less of an issue for GAP design

Outrage ensues

Businessweek: "A Minimum Wage Study with Minimum Credibility"

Political correctness seems to have crept into the inner sanctum of the AEA, discrediting its scholarly journal and debasing its top prize. Unless the association cleans up its act, it can kiss its credibility goodbye

James Buchanan in the WSJ

Just as no physicist would claim that 'water runs uphill,' no self-respecting economist would claim that increases in the minimum wage increase employment. Such a claim, if seriously advanced, becomes equivalent to a denial that there is even minimal scientific content in economics, and that, in consequence, economists can do nothing but write as advocates for ideological interests.

Merton Miller in the WSJ

Raising the minimum wage by law above its market determined equilibrium, they argue, actually costs nobody anything. (Or at worst, costs nobody very much because it's only a small, marginal increment, after all.) Is all this too good to be true? Damn right. But it sure plays well in the opinion polls. I tremble for my profession.

Aftermath

What to make of these results?

- Card-Krueger argue that positive employment effects reflect monopsony power
- Brown (1999) argues that monopsony would imply output expands so prices should fall. Concludes that:

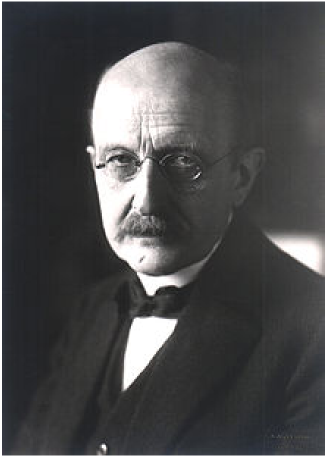
  "Based on the available evidence, the monopsony model will not replace the competitive diagram in the souls of labor economists."

Unresolved: do GAP design and diff in diff identify the same parameter?

- Diff in diff measures market-wide response
- GAP measures effect of raising wage on a single firm holding market constant

One funeral at a time?

A new scientific truth does not triumph by convincing its opponents and making them see the light, but rather because its opponents eventually die, and a new generation grows up that is familiar with it. – Max Planck (1948)



Giuliano (2013)

Study the effect of 1996 fed min wage hike on a large multi-establishment retailer

Leverage high frequency data to assess validity of GAP design

Contrast overall employment effect with relative employment effect (teenage vs adult labor)

Main finding: insignificant aggregate disemployment effect but small *increase* in relative employment of teenage workers

Two Gaps

Gap of employee i at store j is:

$$Gap_{ij} = \max\left\{0, \left(MW_j - w_{ij}\right)/w_{ij}\right\}$$

Store j's average gap is:

$$Gap_j = \frac{1}{N_j}\sum_i Gap_{ij}$$

Store j's relative gap is:

$$\frac{Gap_j^{teen} - Gap_j^{adult}}{1 + Gap_j^{adult}}$$

Assess validity of design via monthly cross sectional regression of outcomes on each gap + controls. Plot coefficients on gap.

Modest effect of avg gap on wage, nothing on employment

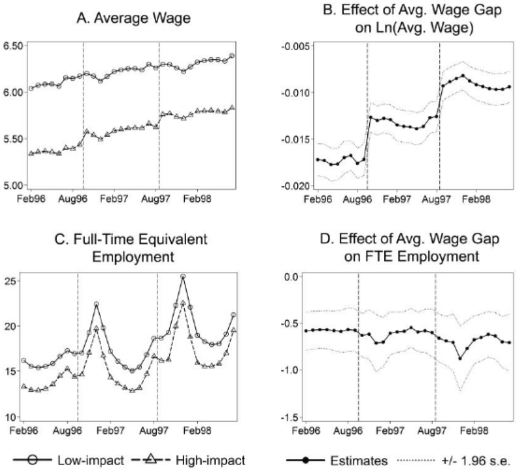
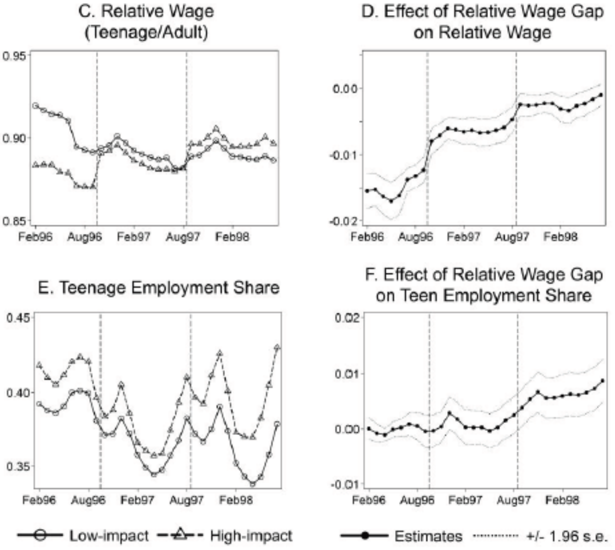


FIG. 1.—Vertical lines indicate dates of federal minimum wage increases (October 1, 1996, and September 1, 1997). Panels *A* and *C* plot group means for high- (low-) impact stores, defined as those with average wage gaps above (below) the sample median. Panels *B* and *D* plot coefficient estimates from monthly regressions of log average wage (*B*) or full-time equivalent employment (*D*) on the store average wage gap (see eq. [2]). Regression models include all fixed store-level controls as in table 4,

Substantial increase in relative wages and employment of teenagers



Quality upgrading?

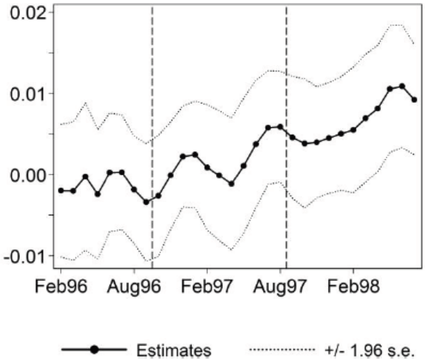


FIG. 4.—Effect of relative wage gap on fraction of teenagers from high-income ZIP codes. Vertical lines indicate dates of federal minimum wage increases (October 1, 1996, and September 1, 1997). Figure plots coefficient estimates from monthly regressions of the fraction of teenage employees who live in high-income ZIP codes on the store relative wage gap (see eq. [2]). Regression model includes all fixed store-level controls as in table 4, col. 4. Coefficients are rescaled to measure differences associated with a .01 difference in the store relative wage gap.

Little evidence of an effect on productivity

Table 10
Reduced-Form Effects of Wage Gaps on Sales Growth and Shrinkage

	Sales Growth			
	(1)	(2)	(3)	(4)
Store adult wage gap	.566^{+}		.639	.911**
	(.296)		(.390)	(.349)
Store teenage wage gap		.194	−.072	−.312
		(.194)	(.257)	(.226)
Sample includes CA, DE, MA, OR, and VT	No	No	No	Yes
Number of stores	>600	>600	>600	>700
R^2	.33	.32	.33	.31
	Change in Shrinkage			
	(1)	(2)	(3)	(4)
Store adult wage gap	−.009		.005	.004
	(.016)		(.019)	(.017)
Store teenage wage gap		−.012	−.014	−.014
		(.012)	(.015)	(.012)
Sample includes CA, DE, MA, OR, and VT	No	No	No	Yes
Number of stores	>600	>600	>600	>700
R^2	.11	.11	.11	.11

NOTE.—Sales growth is the change in the log of total sales between the first 6 months (February–July 1996) and the last 6 months (February–July 1998) of the sample period. Change in shrinkage is the change in the yearly shrinkage rate between the fiscal year ending in February 1996 and the fiscal year ending in February 1998. All regressions include store-level control variables as in table 4, col. 5, plus a control for the change in full-time equivalent employment. Robust standard errors in parentheses.

Note: shrinkage = inventory loss due to shoplifting / theft, etc

Summary

Gap / exposure design seems unconfounded

No discernable effect on overall employment

Relative employment of teens increased slightly

- Many possible explanations: compositional changes, changes in application behavior, monopsony
- Hard to distinguish between them

Limitations:

- Average gap was small
- Difficult to adjust for seasonal in retail employment
- Employment effects might grow over longer horizons..

Harasztosi and Lindner (2019)

US min wage variation tends to be small and short run in nature

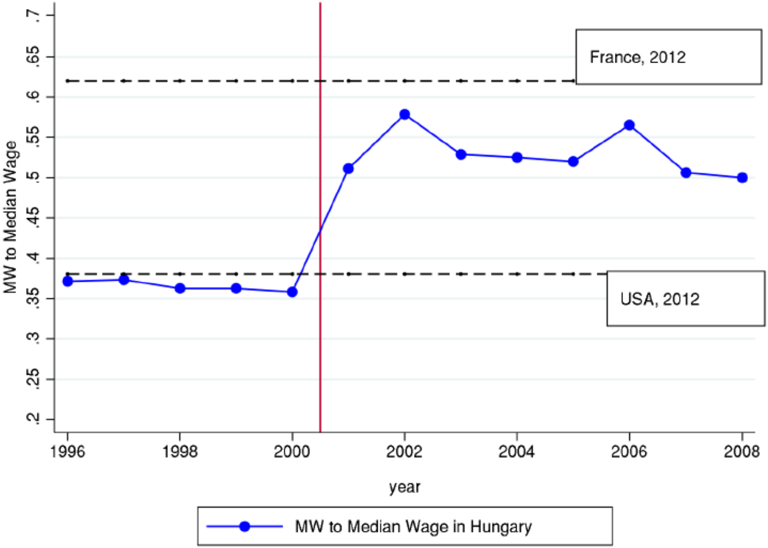
Hungary experienced a large (60%) and persistent ($\sim$8 years) increase in min wage in 2001

Use firm level exposure design to infer MW effects

Findings:

1. Small disemployment effects
2. Substantial cost pass-through to consumers

Huge, permanent, change..



Firm level exposure design

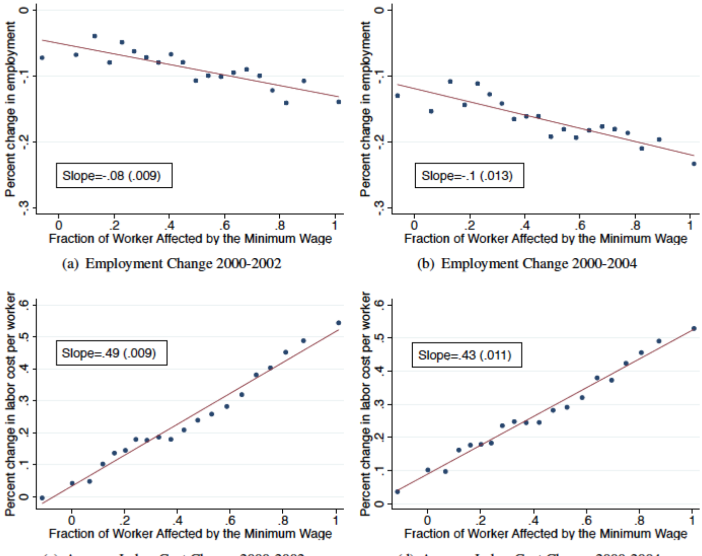
Estimating equations:

$$\frac{y_{it} - y_{i2000}}{y_{i2000}} = \alpha_t + \beta_t FA_{it} + \gamma_t X_{it} + \varepsilon_{it}$$

- y_{it} gives firm i's outcome (employment, wages) in year t
- FA_{it} ("fraction affected") gives the fraction of firm i's employees in 2000 whose wage was below year t minimum
- Weight by log firm revenue in 2000 (logs address extreme skew in revenue)

Firm exposure in 2002 raises wages but lowers employment

FIGURE A4. NON-PARAMETRIC RELATIONSHIP BETWEEN EMPLOYMENT/AVERAGE LABOR COST CHANGE AND THE FRACTION OF AFFECTED WORKERS



(a) Employment Change 2000-2002

(b) Employment Change 2000-2004

(c) Average Labor Cost Change 2000-2002

(d) Average Labor Cost Change 2000-2004

Wage-employment elasticities are small, trivial dynamics

Table 2—Employment and Wage Effects

	Main changes between 2000 and 2002		Main changes between 2000 and 2004		Placebo changes between 2000 and 1998	
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Change in firm-level employment						
Fraction affected	−0.078	−0.076	−0.093	−0.100	−0.003	0.002
	(0.008)	(0.010)	(0.012)	(0.012)	(0.008)	(0.009)
Constant	−0.050		−0.105		0.046	
	(0.005)		(0.007)		(0.005)	
Observations	19,485	19,485	19,485	19,485	19,485	19,485
Employment elasticity with respect to *MW* (directly affected)	−0.11	−0.10	−0.15	−0.15		
	(0.01)	(0.01)	(0.02)	(0.02)		
Panel B. Change in firm-level average wage						
Fraction affected	0.53	0.58	0.48	0.54	−0.02	−0.03
	(0.01)	(0.01)	(0.01)	(0.01)	(0.003)	(0.01)
Constant	0.08		0.16		−0.08	
	(0.002)		(0.01)		(0.001)	
Observations	18,415	18,415	16,980	16,980	19,485	19,485
Employment elasticity with respect to wage	−0.15	−0.13	−0.20	−0.18		
	(0.02)	(0.02)	(0.03)	(0.03)		
Panel C. Change in firm-level average cost of labor						
Fraction affected	0.47	0.49	0.41	0.43	−0.03	−0.04
	(0.01)	(0.01)	(0.01)	(0.01)	(0.003)	(0.01)
Constant	0.04		0.10		−0.04	
	(0.001)		(0.002)		(0.001)	
Observations	18,415	18,415	16,980	16,980	19,485	19,485
Employment elasticity with respect to cost of labor	−0.17	−0.16	−0.22	−0.23		
	(0.02)	(0.02)	(0.03)	(0.03)		
Controls	No	Yes	No	Yes	No	Yes

Notes: This table shows the firm-level relationship between the fraction of workers exposed to the minimum wage and the change in employment (panel A), the change in average wage (panel B), and the change in average cost of labor (panel C). The cost of labor includes wages, social security contributions, and non-wage labor expenses. The

Bias vs Variance

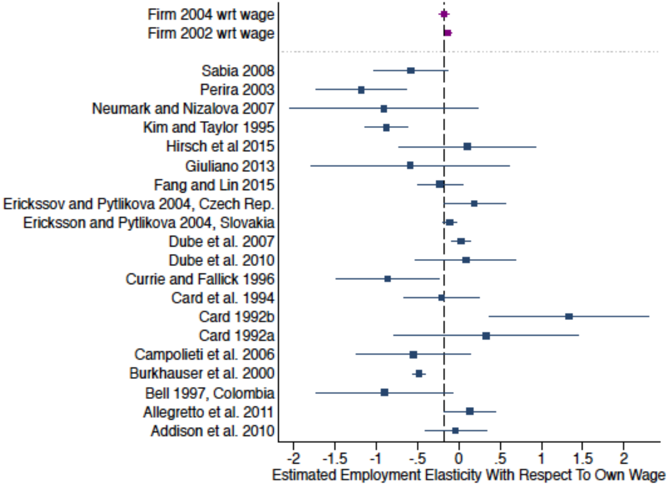
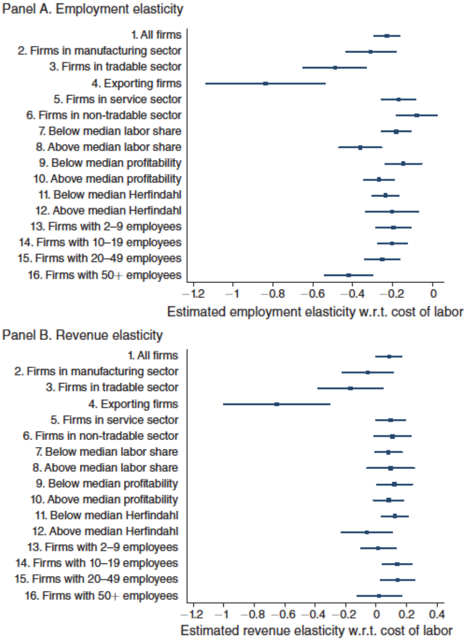


Figure A7. Employment Elasticity in the Literature and in this Paper

Bigger effects in tradeable sectors



Effect on prices

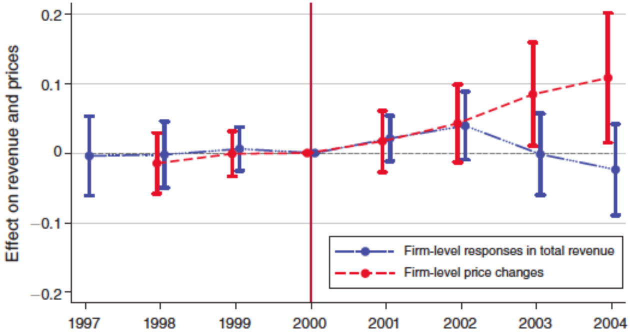


FIGURE 4. EFFECT ON PRICE INDEX AND REVENUE IN THE MANUFACTURING SECTOR

Poor only slightly more likely to consume MW-intensive goods

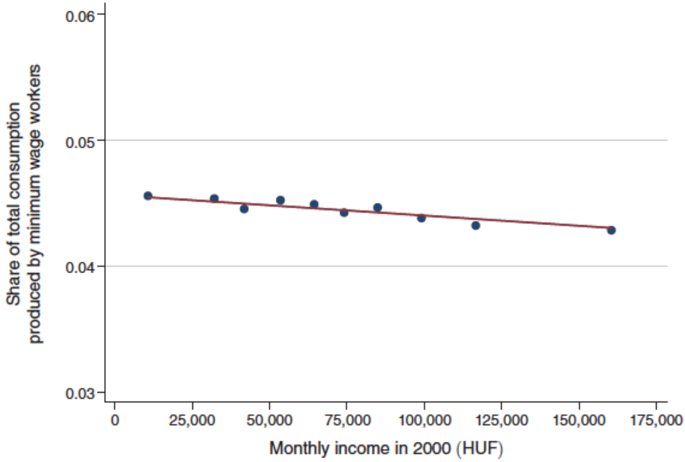


Figure 5. The Relationship between Household Income and the Minimum Wage Content of Consumption

Consumers pay for the min wage

TABLE 5—INCIDENCE OF THE MINIMUM WAGE

	Changes between 2000 and 2002 (1)	Changes between 2000 and 2004 (2)
Change in total labor cost relative to revenue in 2000	0.038	0.021
Change in revenue relative to revenue in 2000(Δ *Revenue*)	0.066	0.036
Change in materials relative to revenue in 2000 (Δ *Material*)	0.033	0.014
Change in miscitems relative to revenue in 2000 (Δ *MiscItems*)	0.005	0.005
Incidence on consumers (Δ *Revenue* $-$ Δ *Material* $-$ Δ *MiscItems*)	0.028	0.017
Change in profits relative to revenue in 2000 (Δ *Profit*)	−0.011	−0.008
Change in depreciation relative to revenue in 2000 (Δ *Depr*)	0.001	0.003
Incidence on firm owners ($-\Delta$ *Profit* $-$ Δ *Depr*)	0.010	0.005
Fraction paid by consumers (percent)	74	77
Fraction paid by firm owners (percent)	26	23

A rationalizing framework

- Monopsonistic competition: each firm produces a different product variety ω
- Three factors of production: labor, capital, materials. Derived labor demand is $l(\omega)$
- Model yields *firm*-level demand elasticities

$$\frac{\partial \ln l(\omega)}{\partial \ln MW} = \underbrace{-s_L\eta}_{\text{scale effect}} - \underbrace{s_K\sigma_{KL}}_{\text{K-L substitution}} - \underbrace{s_M\sigma_{ML}}_{\text{ML substitution}}$$

$$\frac{\partial \ln p(\omega)\, q(\omega)}{\partial \ln MW} = \underbrace{s_L}_{\text{price effect}} - \underbrace{s_L\eta}_{\text{scale effect}}$$

$$\frac{\partial \ln k(\omega)}{\partial \ln MW} = s_L(-\eta + \sigma_{KL}), \frac{\partial \ln m(\omega)}{\partial \ln MW} = s_L(-\eta + \sigma_{ML})$$

Note: η is determined in *equilibrium* and depends on fraction of firms affected by min wage. It is smallest when all firms are affected by min wage.

Estimation

Calibrate shares (s_L, s_K, s_M) from microdata leaving 3 unknown structural elasticities:

$$\eta, \sigma_{KL}, \sigma_{ML}$$

- Proxy elasticity wrt MW with treatment effects
  - Recall that MW change is large, so implicitly assuming iso-elastic demand
  - 4 equations and 3 unknowns $\Rightarrow$ over-determined system
- Fit via classical minimum distance (see Wooldridge, 2010)
  - Equivalent to stacking moments and treating as an SUR

SUR representation

Dataset of four moments, 3 regressors (s_L,s_K,s_M), no intercept:

$$
\begin{aligned}
\frac{\partial \widehat{\ln l(\omega)}}{\partial \ln MW} &= -\eta \cdot s_L - \sigma_{KL} \cdot s_K - \sigma_{ML} \cdot s_M + \varepsilon_L \\
\frac{\partial \ln \widehat{p(\omega)} \, q(\omega)}{\partial \ln MW} &= (1-\eta) \cdot s_L - 0 \cdot s_K - 0 \cdot s_M + \varepsilon_R \\
\frac{\partial \widehat{\ln k(\omega)}}{\partial \ln MW} &= (\sigma_{KL} - \eta) \cdot s_L - 0 \cdot s_K - 0 \cdot s_M + \varepsilon_K \\
\frac{\partial \widehat{\ln m(\omega)}}{\partial \ln MW} &= (\sigma_{ML} - \eta) \cdot s_L - 0 \cdot s_K - 0 \cdot s_M + \varepsilon_M
\end{aligned}
$$

where $Cov(\varepsilon_L, \varepsilon_R, \varepsilon_K, \varepsilon_M) = \Sigma$ is estimated from microdata

- Here SUR $=$ multivariate weighted least squares
- Coefficient restrictions exploited in conjunction w/ Σ to improve precision

Materials key to getting neoclassical model to work..

TABLE 6—METHOD OF THE MOMENTS ESTIMATES USING MEDIUM-TERM RESPONSES

	All firms (1)	Manufacturing (2)	Tradable (3)	Non-tradable (4)	Export (5)
Panel A. Estimated parameters					
Output demand, η	0.11	0.98	1.34	−0.37	3.64
	(0.22)	(0.46)	(0.41)	(0.50)	(0.98)
Capital-labor substitution, σ_{KL}	3.35	2.60	2.34	3.94	4.63
	(0.62)	(1.01)	(0.83)	(1.59)	(2.45)
Material-labor substitution, σ_{ML}	0.03	0	0.01	0	0
	(0.06)	(0.10)	(0.13)	(0.09)	(0.26)
Panel B. Empirical moments					
Employment elasticity	−0.23	−0.31	−0.49	−0.08	−0.84
Revenue elasticity	0.08	−0.05	−0.17	0.11	−0.65
Materials elasticity	0.05	−0.17	−0.26	0.04	−0.73
Capital elasticity	0.62	0.37	0.28	0.70	0.50
Price elasticity		0.25			
Panel C. Moments predicted by the estimated parameters					
Employment elasticity	−0.24	−0.33	−0.51	−0.12	−0.95
Revenue elasticity	0.16	0.003	−0.09	0.12	−0.49
Materials elasticity	−0.01	−0.18	−0.33	0	−0.67
Capital elasticity	0.58	0.29	0.23	0.22	0.1
Price elasticity	0.18	0.23	0.25	0.12	0.18
Share of labor, s_L	0.18	0.23	0.25	0.12	0.18
Share of capital, s_K	0.08	0.07	0.08	0.07	0.08
Share of materials, s_M	0.74	0.70	0.67	0.81	0.74
No. of moments used	4	4	4	4	4
No. of estimated parameters	3	3	3	3	3
SSE	5.64	0.76	1.00	2.20	2.02

Notes: We estimate the parameters of the model presented in Section V using a minimum-distance estimator. In each column we use the empirical moments based on our benchmark estimates with controls. The estimated parameters with standard errors can be found in panel A. Panels B and C report the empirical and the predicted moments, respectively. SSE reports the weighted sum of squared errors.

Summary

Cost effects of min wage largely passed through to consumers!
(cost-push inflation / redist thru prod market)
Materials share key to rationalizing small losses

Summary

Cost effects of min wage largely passed through to consumers!
(cost-push inflation / redist thru prod market)
Materials share key to rationalizing small losses

Policy implications

- Price increases in tradeable sectors are a win for small country like Hungary
- Price increases to domestic consumers more problematic

Summary

Cost effects of min wage largely passed through to consumers!
(cost-push inflation / redist thru prod market)
Materials share key to rationalizing small losses

Policy implications

- Price increases in tradeable sectors are a win for small country like Hungary
- Price increases to domestic consumers more problematic

Generalizability:

- What about the U.S.? What are the effects of smaller less persistent MW changes?
- How to distinguish effects of *firm*-specific from aggregate MW changes?

Cengiz, Dube, Lindner, and Zipperer (2019)

Examine 138 state minimum wage changes in the U.S.

Assess impact on state-wide frequency distribution of wages via DiD

- Publicly available data! CPS benchmarked to QCEW

Methodological insight: use distributional impacts to infer employment losses

- How does this work?
- Recall: impact on distribution $\neq$ distribution of impacts!

The basic idea

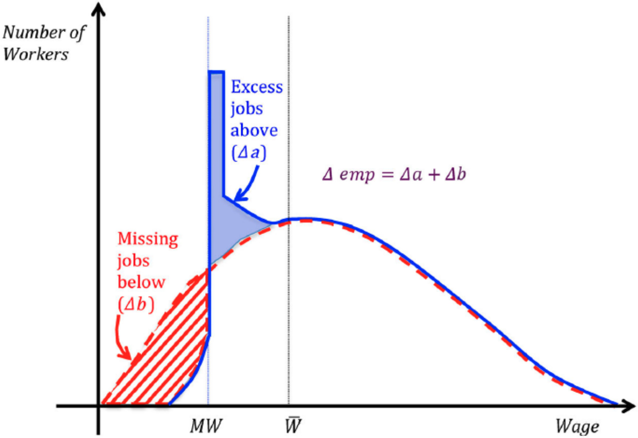


FIGURE I

The Impact of Minimum Wages on the Frequency Distribution of Wages

Key assumptions: exclusion restriction (no effect above $\bar{W}$) + sign restrictions (emp gains in $\left[MW, \bar{W}\right]$, losses in $(0, MW)$)

Estimating job loss

Distributional event study specification:

$$\frac{E_{sjt}}{N_{st}} = \sum_{\tau=-3}^{4} \sum_{k=-4}^{17} \alpha_{\tau k} I_{sjt}^{\tau k} + \mu_{sj} + \rho_{jt} + \Omega_{sjt} + u_{sjt}$$

- E_{sjt} is employment in \$0.25 wage bin j of state s at time t
- N_{st} is population in state s at time t
- $\alpha_{\tau k}$ effect of min wage hike τ periods ago on bins $\$[k, k+1]$ above state min wage
- μ_{sj} state by wage bin FE
- ρ_{jt} bin by year FE
- Ω_{sjt} controls for "small" min wage changes
- Cluster on state (i.e., meta-analysis std errors)

Target parameters

Distributional event study specification:

$$\frac{E_{sjt}}{N_{st}} = \sum_{\tau=-3}^{4} \sum_{k=-4}^{17} \alpha_{\tau k} I_{sjt}^{\tau k} + \mu_{sj} + \rho_{jt} + \Omega_{sjt} + u_{sjt}$$

- Scaled decrease in employment below new minimum

$$\Delta a_\tau = \frac{\sum_{k=-3}^{-1} (\alpha_{\tau k} - \alpha_{-1k})}{EPOP_{-1}}$$

- Scaled increase above new minimum (setting $\bar{W} - MW = 4$) due to "bunching":

$$\Delta b_\tau = \frac{\sum_{k=0}^{4} (\alpha_{\tau k} - \alpha_{-1k})}{EPOP_{-1}}$$

- Net (scaled) employment change at horizon τ is $\Delta a_\tau + \Delta b_\tau$

No net employment losses 5 years out..

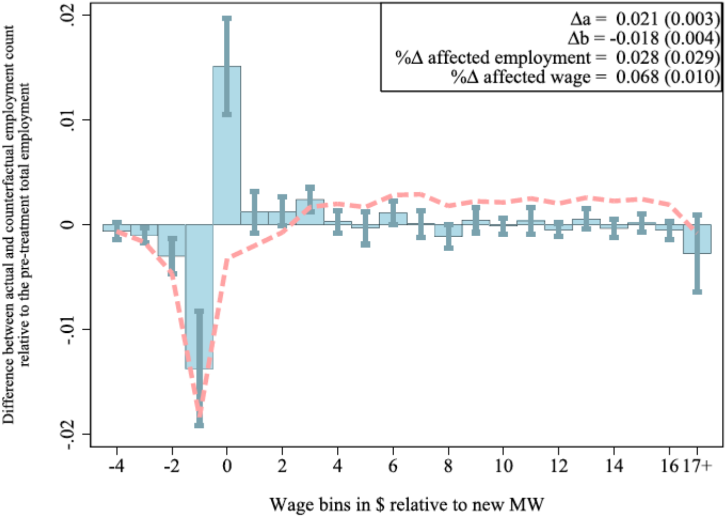


FIGURE II

Impact of Minimum Wages on the Wage Distribution

Not much in the way of dynamics

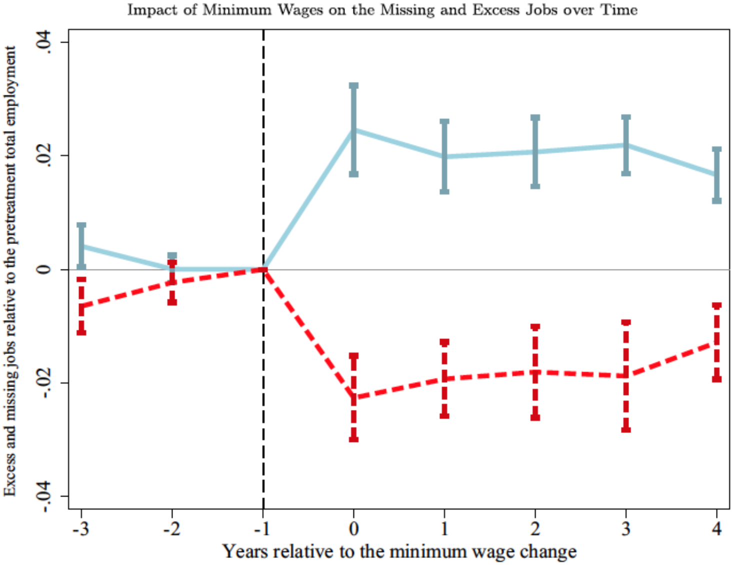


FIGURE III

Impact of Minimum Wages on the Missing and Excess Jobs over Time

No net employment effect on new entrants or incumbents

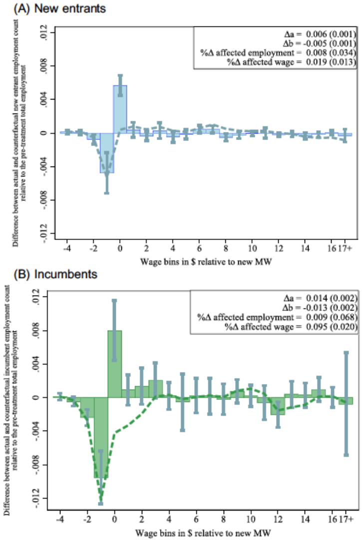


FIGURE IV

Impact of Minimum Wages on the Wage Distribution by Pretreatment Employment Status: New Entrants and Incumbents

Substantial wage spillovers above minimum

Except among new entrants and tradeable sectors..

TABLE IV
THE SIZE OF THE WAGE SPILLOVERS

	%Δ Affected wage		Spillover share of wage increase
	$\%\Delta w$	$\%\Delta w_{\text{No spillover}}$	$\frac{\%\Delta w - \%\Delta w_{\text{No spillover}}}{\%\Delta w}$
Overall	0.068***	0.041***	0.397***
	(0.010)	(0.009)	(0.119)
Less than high school	0.077***	0.048***	0.370***
	(0.013)	(0.009)	(0.078)
Teen	0.081***	0.053***	0.347***
	(0.015)	(0.007)	(0.059)
High school or less	0.073***	0.043***	0.402***
	(0.013)	(0.011)	(0.100)
Women	0.070***	0.045***	0.359***
	(0.011)	(0.010)	(0.120)
Black or Hispanic	0.045***	0.037***	0.179
	(0.012)	(0.010)	(0.265)
Tradeable	0.058	0.065**	−0.114
	(0.073)	(0.028)	(1.157)
Nontradeable	0.056***	0.043***	0.237
	(0.014)	(0.006)	(0.191)
Incumbent	0.095***	0.055***	0.422**
	(0.020)	(0.011)	(0.181)
New entrant	0.019	0.023***	−0.178
	(0.013)	(0.006)	(0.748)

Notes. The table reports the effects of a minimum wage increase on wages based on the event study analysis (see equation (1)) exploiting 138 state-level minimum wage changes between 1979 and 2016. The table reports the percentage change in affected wages with (column (1)) and without (column (2)) taking spillovers into account for all workers, workers without a high school degree, teens, individuals with high school or less schooling, women, black or Hispanic workers, in tradeable industries, in nontradeable industries, those who were employed one year before the minimum wage increase (incumbents); and those who did not have a job one year before (new entrants). The first column is the estimated change in the affected wages calculated according to equation (2), and the second column assumes no spillovers (see equation (3)). In the last column, the spillover share of the wage effect is calculated by subtracting 1 from the ratio of the estimates in the second to the first column. Robust standard errors in parentheses are clustered by state; significance levels are *0.10, **0.05, ***0.01.

Reconciling with conventional panel estimates

The behavior of the mean is sensitive to the response of very high wages

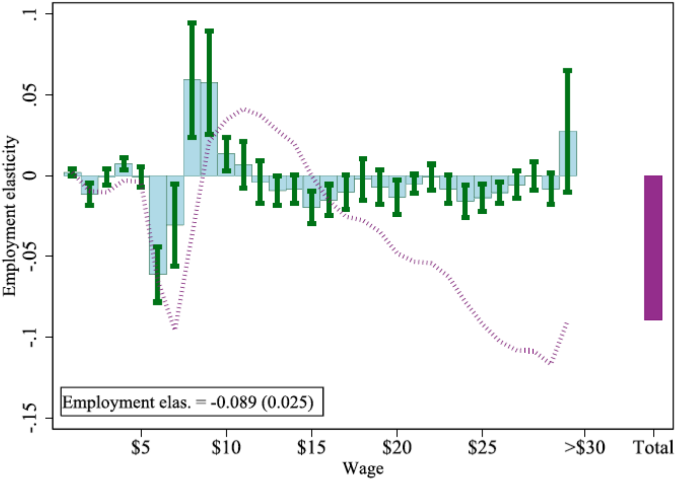


FIGURE VI

Impact on Employment throughout the Wage Distribution in the Two-Way Fixed Effects Model on Log Minimum Wages

Summary

Market-wide fluctuations in state min wages seem to generate tiny employment losses or even small employment gains

No appreciable dynamics

Prior time series analyses of aggregates likely confounded by sensitivity of mean to top quantiles

Methodological lesson: power of going beyond the mean

Key insight of paper was that MW fluctuations should not strongly affect top quantiles of the wage distribution $\Rightarrow$ sufficient to evalute MW impact on jobs with $W < \bar{W}$

- Used distributional regressions to find threshold $\bar{W}$ above which treatment effects are trivial
- See Fortin, Lemieux, and Lloyd (2018) for other approaches to distributional regression

Implicitly a restriction on *joint distribution* of potential outcomes: workers pushed from wages levels below to just above new MW

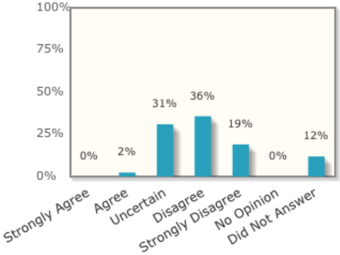
- Reforms only affecting attractiveness of some options and not others often yield similar identification of adjustment margins
- Papers estimating "flows" between counterfactual choices: Kline and Tartari (2016), Kline and Walters (2016), Pinto (2018)

Little support in 2015 for view that MW increases productivity

Question B:

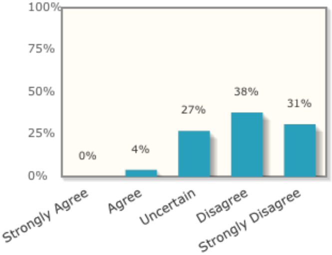
Increasing the federal minimum wage gradually to $15-per-hour by 2020 would substantially increase aggregate output in the US economy.

Responses



Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Responses weighted by each expert's confidence



Source: IGM Economic Experts Panel
www.igmchicago.org/igm-economic-experts-panel

Dustmann, Lindner, Schoenberg, Umkehrer, vom Berge (2019)

- Min wage lit has focused on market level tradeoff between employment and wage inequality
- But in principle there can also be a beneficial *reallocation* effect of the minimum wage: workers move from less to more productive firms.
- Can potentially raise total productivity in the economy if market imperfections "protected" unproductive firms in the first place (Hsieh and Klenow, 2009) or if capital intensity was inefficiently low to begin with (Acemoglu, 2001).

- Germany instituted first national minimum wage in January 2015
  - NOT indexed to local cost of living so disproportionately hit less productive East German firms
  - Study reallocation effect of policy

Distributional DiD (initially low vs high wage workers)

Wages of affected individuals go up, no effect on employment probs

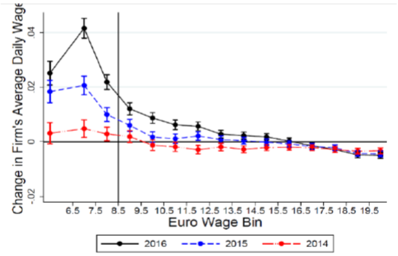
Table 2: Effect of the Minimum Wage on Wages and Employment: Individual Approach

	(1)	(2)	(3)	(4)	(5)
	Changes relative to 2011 vs 2013			Difference-in-difference	
Wage bin in *t-2*	[4.5,8.5)	[8.5,12.5)	[12.5,20.5)	(1) minus (3)	(2) minus (3)
Panel (a): Hourly Wages					
2014 vs 2016	0.067	0.023	0.006	0.061	0.016
	(0.0006)	(0.0003)	(0.0001)	(0.0006)	(0.0003)
2012 vs 2014 (Placebo)	0.017	0.009	0.006	0.010	0.003
	(0.0005)	(0.0003)	(0.0001)	(0.0006)	(0.0003)
Baseline Change (2011 vs 2013)	0.199	0.118	0.080		
Panel (b): Daily Wages					
2014 vs 2016	0.118	0.047	0.012	0.107	0.036
	(0.0010)	(0.0005)	(0.0002)	(0.0010)	(0.0005)
2012 vs 2014 (Placebo)	0.022	0.012	0.006	0.015	0.006
	(0.0009)	(0.0005)	(0.0002)	(0.0009)	(0.0005)
Baseline Change (2011 vs 2013)	0.220	0.064	-0.002		
Panel (c): Employment (1 if employed)					
2014 vs 2016	0.009	0.003	0.002	0.007	0.001
	(0.0004)	(0.0002)	(0.0001)	(0.0004)	(0.0003)
2012 vs 2014 (Placebo)	0.003	0.000	0.001	0.002	-0.001
	(0.0004)	(0.0002)	(0.0001)	(0.0004)	(0.0003)
Baseline Change (2011 vs 2013)	-0.242	-0.184	-0.141		
Panel (d): Employment, full-time equivalents					
2014 vs 2016	0.034	0.018	0.006	0.029	0.013
	(0.0004)	(0.0002)	(0.0001)	(0.0004)	(0.0003)
2012 vs 2014 (Placebo)	0.010	0.006	0.002	0.009	0.004
	(0.0003)	(0.0002)	(0.0001)	(0.0004)	(0.0003)
Baseline Change (2011 vs 2013)	-0.180	-0.193	-0.179		

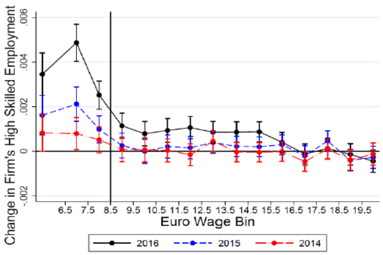
Notes: In panel (a), we report the excess hourly wage growth in the 2014 vs 2016 post-policy period and

Initially low wage workers move to better firms

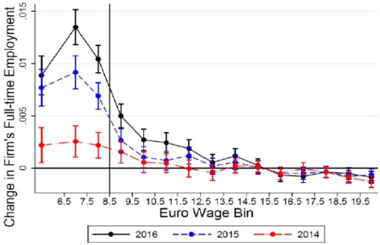
Figure 4: Reallocation Effects of the Minimum Wage: Individual Approach



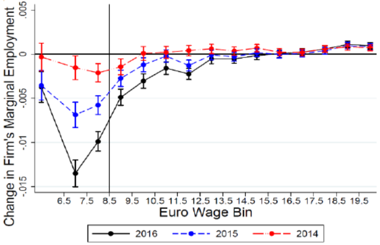
(a) Firm's Average Daily Wage



(b) Firm's High-Skilled Employment Share



(c) Firm's Full-Time Employment Share



(d) Firm's Marginal Employment Share

Higher firm wage FEs and lower churn rates

Figure 5: Reallocation Effects of the Minimum Wage: Individual Approach

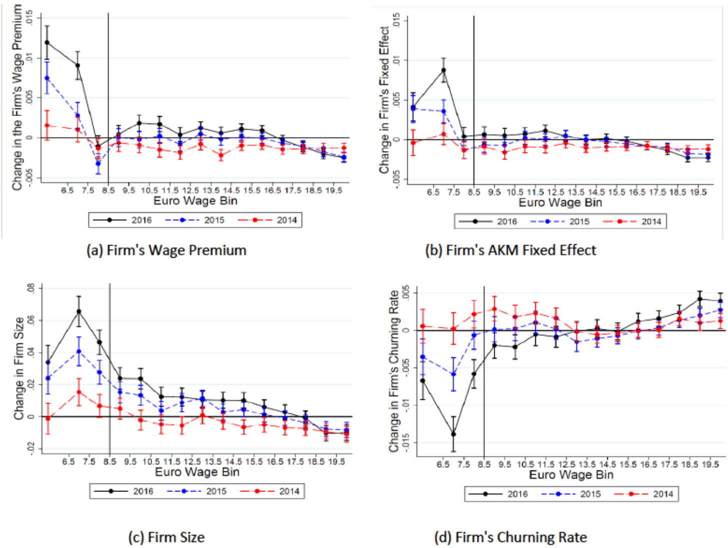
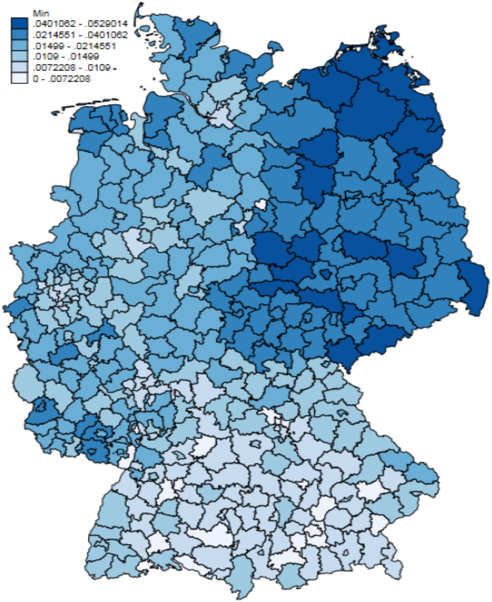


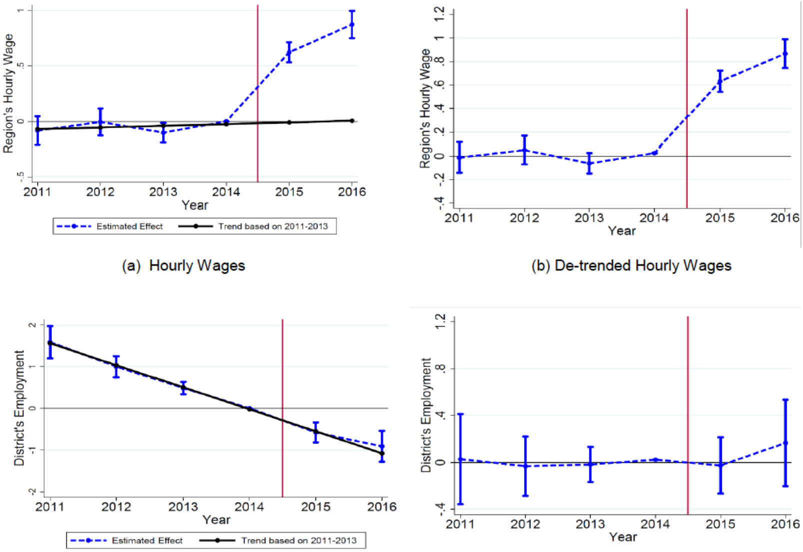
Figure 5: Exposure to the Minimum Wage across Regions



Notes: The figure shows the exposure to the minimum wage across 401 regions (districts). Regional exposure to the minimum wage is measured using the gap

Big effects on *market* wages, nothing on employment

Figure 7 : Wage and Employment Effects of the Minimum Wage: Regional Approach



(a) Hourly Wages

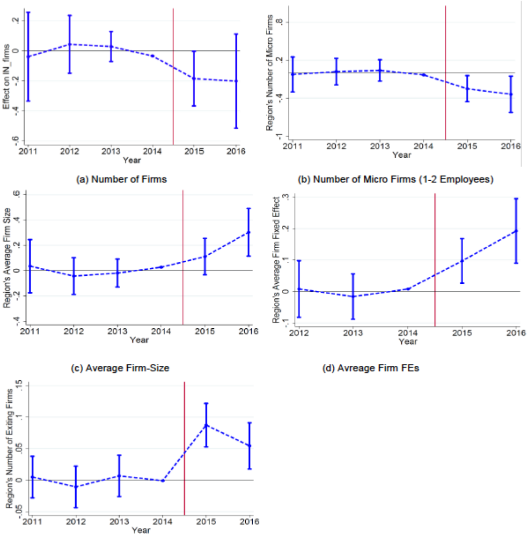
(b) De-trended Hourly Wages

(c) Employment

(d) De-trended Employment

Reallocation from smaller to bigger firms

Figure 8: Evidence for Reallocation: Regional Aproach



(a) Number of Firms

(b) Number of Micro Firms (1-2 Employees)

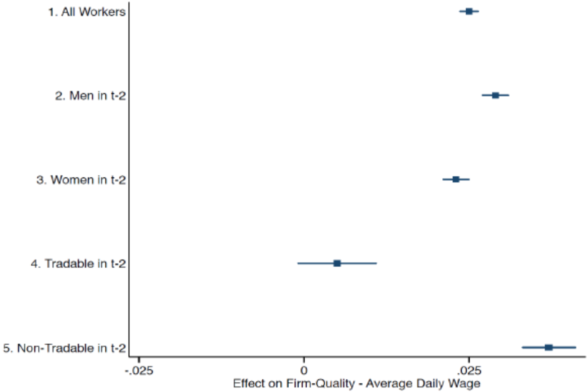
(c) Average Firm-Size

(d) Avreage Firm FEs

(e) Firm Exit

Bigger effects among workers initially in non-tradeables

Figure 9: Heterogeneity of Reallocation Responses



Notes: This figure shows the effect of the minimum wage on the reallocation of low-wage workers to firms that pay a higher average daily wages. Row (1) shows the benchmark estimate when all workers are included in the sample (as in panel (b) in Table 3). In rows (2) and (3), the sample is split into men and women, respectively. Rows (4) and Row (5) estimate the reallocation effect separately for workers who were employed in the tradable and in the non-tradable sector at baseline. We classify sectors into tradable and non-tradable using method 1 in Mian and Sufi (2014).

Summary

- Labor markets are frictional and, when left to their own, can generate misallocation
- The minimum wage seems to kill less productive firms in less competitive industries
- But no effect on aggregate employment because workers are reallocated to more productive businesses
- Possible that total output rose (allocative efficiency)

References

1. Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2020). Sampling-Based versus Design-Based Uncertainty in Regression Analysis. Econometrica, 88(1), 265-296.
2. Acemoglu, D. (2001). Good jobs versus bad jobs. Journal of labor Economics, 19(1), 1-21.
3. Andrews, I., & Kasy, M. (2019). Identification of and correction for publication bias. American Economic Review, 109(8), 2766-94.
4. Brown, C. (1999). Minimum wages, employment, and the distribution of income. Handbook of labor economics, 3, 2101-2163.
5. Card, D., & Krueger, A. B. (1994). Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania. American Economic Review, 84, 772-793.
6. Card, D., & Krueger, A. B. (1995). Time-series minimum-wage studies: a meta-analysis. The American Economic Review, 85(2), 238-243.
7. Cengiz, D., Dube, A., Lindner, A., & Zipperer, B. (2019). The effect of minimum wages on low-wage jobs. The Quarterly Journal of Economics, 134(3), 1405-1454.

8. Dustmann, C., Lindner, A., Schonberg, U., Umkehrer, M., and vom Berge, P. (2019) "Reallocation Effects of the Minimum Wage: Evidence From Germany" working paper.
9. Giuliano, L. (2013). Minimum wage effects on employment, substitution, and the teenage labor supply: Evidence from personnel data. Journal of Labor Economics, 31(1), 155-194.
10. Harasztosi, P., & Lindner, A. (2019). Who Pays for the minimum Wage?. American Economic Review, 109(8), 2693-2727.
11. Hsieh, C. T., & Klenow, P. J. (2009). Misallocation and manufacturing TFP in China and India. The Quarterly journal of economics, 124(4), 1403-1448.
12. Fortin, N., Lemieux, T., & Lloyd, N. (2018). Labor market institutions and the distribution of wages: the role of spillover effects.
13. Kline, P., & Tartari, M. (2016). Bounding the labor supply responses to a randomized welfare experiment: A revealed preference approach. American Economic Review, 106(4), 972-1014.
14. Kline, P., & Walters, C. R. (2016). Evaluating public programs with close substitutes: The case of Head Start. The Quarterly Journal of Economics, 131(4), 1795-1848.
15. Pinto, R. (2018). Noncompliance as a rational choice: A framework that exploits compromises in social experiments to identify causal effects.

16. Stigler, G. J. (1946). The economics of minimum wage legislation. The American Economic Review, 36(3), 358-365.
17. Wooldridge, J. M. (2010). Econometric analysis of cross section and panel data. MIT press.

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture #2: Firm Wage Premia

Patrick Kline

UC Berkeley

Slichter (1950): a 1940 wage survey from Boston

	Average Hourly Earnings in All Plants (cents)	Average Hourly Earnings in Lowest Plant (cents)	Average Hourly Earnings in Highest Plant (cents)	Spread between High and Low Plants (cents)
Common labor	57.9	44.8	74.1	29.3
Janitor	55.3	41.0	70.5	29.5
Watchman	59.6	45.2	74.0	28.8
Producing and processing laborers	64.2	44.8	100.7	55.9
Producing and processing operators	72.0	57.8	88.6	30.8
Receiving and shipping clerks	68.0	50.0	89.6	39.6
Machinists	87.5	70.0	105.0	35.0
Steamfitter	86.4	70.0	105.0	35.0
Electrician	88.0	67.9	105.0	37.1
Carpenter	81.5	65.0	99.5	34.5
Sheet metal workers	85.4	77.8	90.5	12.7
Millwright	86.1	82.5	95.5	13.0
Maintenance helper	67.1	50.7	82.0	31.3
Female producing laborers	45.1	33.8	63.4	29.6
Female producing operators	47.9	37.7	58.3	20.6
Firemen	78.4	63.0	90.8	27.8

Slichter: "neither wage rates nor hourly earnings represent the price of labor"

- Slichter studies "structure of wages" using industry-level data from 1939 Economic Census (firm data was not available)
- Discovers 7 regularities about wages of unskilled men:

1. Positive correlation with wages of skilled co-workers
2. Negative correlation with % female
3. Positive correlation with industry value-added / worker-hour
4. Positive correlation with sales / worker-hour
5. Negative correlation with payroll / sales
6. Positive correlation with sales margin (i.e. value added / sales)
7. Stable over time (high correlation of industry wage rank)

- Interpretation: "the results of this study give strong support to the proposition that managerial policy is important in determining inter-industry wage differences."

Krueger and Summers (1988)

- Was Slichter right that some industries pay higher wages?
- Use panel data to study what happens when workers switch industries
- Compare to cross-sectional estimates of wage premia to infer degree of unobserved sorting

Bias correcting the variance of fixed effect estimates

All models amount to:

$$y_i = D_i'\beta + X_i\delta + \varepsilon_i$$

where D_i is a vector of industry dummies and X_i is a vector of controls that may or may not include individual fixed effects

- If each industry fixed effect $\hat{\beta}_j \sim N\left(\beta_j, \sigma_j^2\right)$, then

$$\mathbb{E}\left[\hat{\beta}_j^2\right] = \beta_j^2 + \sigma_j^2.$$

- Suppose we have consistent standard error estimates $\{\hat{\sigma}_j\}_{j=1}^J$
- Then a consistent bias corrected standard deviation of industry wage premia is

$$\sqrt{\frac{1}{J-1}\sum_{j=1}^{J}\left(\hat{\beta}_j - \bar{\beta}\right)^2 - \frac{1}{J}\sum_{j=1}^{J}\hat{\sigma}_j^2}$$

Note: ignoring variability in $\bar{\beta} = \frac{1}{J}\sum_{j=1}^{J}\beta_j$ which is of smaller order.

Substantial cross sectional variability

TABLE I

ESTIMATED WAGE DIFFERENTIALS FOR ONE-DIGIT INDUSTRIES—MAY CPS[a]
(Standard Errors in Parentheses)

Industry	(1) 1974	(2) 1979	(3) 1984	(4) 1984 Total Compensation
Construction	.195	.126	.108	.091
	(.021)	(.031)	(.034)	(.035)
Manufacturing	.055	.044	.091	.131
	(.020)	(.029)	(.032)	(.032)
Transportation & Public Utilities	.111	.081	.145	.203
	(.021)	(.031)	(.034)	(.034)
Wholesale & Retail Trade	−.128	−.082	−.111	−.136
	(.020)	(.030)	(.033)	(.033)
Finance, Insurance and Real Estate	.047	−.010	.055	.069
	(.022)	(.035)	(.034)	(.034)
Services	−.070	−.055	−.078	−.111
	(.021)	(.030)	(.032)	(.032)
Mining	.179	.229	.222	.231
	(.035)	(.058)	(.075)	(.075)
Weighted Adjusted Standard Deviation of Differentials[b]	.097**	.069**	.094**	.126**
Sample Size	29,945	8,978	11,512	11,512

[a] Other explanatory variables are education and its square, 6 age dummies, 8 occupation dummies, 3 region dummies, sex dummy, race dummy, central city dummy, union member dummy, ever married dummy, veteran status, marriage × sex interaction, education × sex interaction, education squared × sex interaction, 6 age × sex interactions, and a constant. Each column was estimated from a separate cross-sectional regression.

[b] Weights are employment shares for each year.

** F test that industry wage differentials jointly equal 0 rejects at the .000001 level.

Worker FE estimates ≈ Cross-Sectional Estimates!

TABLE IV

THE EFFECTS OF UNMEASURED LABOR QUALITY[a]

Industry	(1) Fixed Effects Unadjusted for Measurement Error	(2) Fixed Effects Adjusted for Measurement Error I[b]	(3) Fixed Effects Adjusted for Measurement Error II[c]	(4) Levels
Construction	.063	.098	.174	.174
	(.033)	(.060)	(.060)	(.024)
Manufacturing	.028	.055	.107	.064
	(.031)	(.058)	(.058)	(.022)
Transportation and Public Utilities	.019	.060	.049	.114
	(.035)	(.059)	(.059)	(.024)
Wholesale and Retail Trade	−.042	−.068	−.125	−.133
	(.031)	(.056)	(.056)	(.023)
Finance, Insurance and Real Estate	.027	.017	.018	.035
	(.036)	(.061)	(.061)	(.025)
Services	−.040	−.088	−.128	−.079
	(.032)	(.056)	(.057)	(.023)
Mining	.067	.122	.142	.156
	(.004)	(.057)	(.058)	(.040)

[a] Data set is three matched May CPS's pooled together: 1974–1975, 1977–1978, and 1979–1980. Sample size is 18,122. Levels are 1974, 1977, and 1979 data pooled. Results of the 1975, 1978, and 1980 sample are qualitatively the same. Controls for fixed effects regressions are change in education and its square, change in occupation, 3 region dummies, change in union membership, experience squared, change in marital status, year dummies, and a constant. Controls for level regressions are the same as Table I plus year dummies.

[b] Adjustment I assumes 3.4 per cent error rate and that misclassifications are proportional to industry size. See Appendix for description.

[c] Adjustment II assumes average error rate is 3.4 per cent and misclassifications are allocated according to employer-employee mismatches. See Appendix for description.

No evidence of compensating differentials

TABLE VI

ANALYSIS OF INDUSTRY WAGE DIFFERENTIALS WITH AND WITHOUT CONTROLS FOR WORKING CONDITIONS—QES 1977[a]

Industry	Coefficient (SE) (1)	Coefficient (SE) (2)
Construction	.113	.100
	(.098)	(.100)
Manufacturing	.050	.046
	(.086)	(.087)
Transportation	.113	.124
	(.095)	(.096)
Wholesale & Retail Trade	−.056	−.061
	(.090)	(.091)
Finance, Insurance and	.071	.053
Real Estate	(.104)	(.105)
Services	−.107	−.104
	(.090)	(.091)
Mining	.233	.308
	(.205)	(.220)
10 Working Condition Variables[b]	no	yes
Weighted Adjusted Standard Deviation of 2-Digit Industry Premiums	.113*	.118*
R^2	.496	.519

[a] Other explanatory variables are education and its square, derived experience and its square, sex, race, 3 region dummies, tenure with employer and its square, union status, and 8 occupation dummies. Sample size is 1,033.

[b] Working condition variables are weekly hours, variables indicating dangerous or unhealthy conditions on the job and whether the danger/threat is serious, commuting time, second and third shift dummies, two dummies indicating extent of choice of overtime, and two dummies indicating whether the physical working conditions are pleasant.

* F test that industry wage differentials jointly equal 0 is rejected at .00005 level.

People don't quit high wage jobs

TABLE IX

THE EFFECT OF INDUSTRY WAGE DIFFERENTIALS ON JOB TENURE AND QUITS

Independent Variables	Dependent Variable[a] (1) Tenure	(2) Quit[b]
Industry wage premium	2.198 (.676)	−.073 (.135)
Union (1 = yes)	3.179 (.157)	−.164 (.037)
Other variables	Age dummies (6), Age ∗ Sex (6), Education, Education Squared ∗ Sex, Region Dummies (3), Race Dummy, Sex Dummy, Central City Dummy, Firm Size Dummies (4), Plant Size Dummies (4), Marriage Dummy, Marriage ∗ Sex, Veteran Status Dummy	Education, Education Squared, Region Dummies (3), Race Dummy, Sex Dummy, SMSA Dummy, (Age—Education—5) and its square
Sample Size	8,978	633
R^2	.40	.20

[a] Mean (SD) of Tenure is 5.70 (7.61); Mean (SD) of Quit is .26 (.44).
[b] Quit equation was estimated with a linear probability model.

Gibbons and Katz (1992)

Even first differenced estimates of industry wage premia biased if there is sorting based on match effects

Basic idea:

- Good workers work in more productive industries but would be paid the same amount everywhere if known to be good
- When a worker is revealed to be “good” she moves to the good industry and experiences a wage change
- But the causal mechanism is the revelation that she is good, not the industry

Test by looking at exogenous separations associated with plant closings (as measured in CPS Displaced Workers Survey)

Mild evidence of endogenous mobility

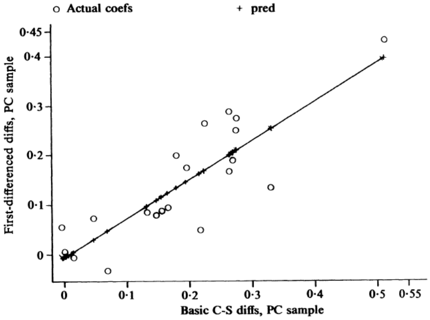


FIGURE 1
First-differenced vs. C − S ind diffs, PC sample

Industry premia estimates from switchers very close to cross-sectional:

- Plant closing sample (pictured above): Slope=0.79, $R^2 = 0.72$
- Layoffs sample (not pictured): Slope = .91, $R^2 = 0.81$

Suggests variance of industry wage effects somewhat overstated due to endogenous mobility

- Perhaps also treatment effect heterogeneity?

Abowd, Kramarz, Margolis (1999)

Industry is just a linear combination of firms. Are there big pay differences within the same industry?

- Use Employer-Employee data to study firm switchers
- Fixed effects specification:

$$y_{it} = \underbrace{\alpha_i}_{\text{worker eff}} + \underbrace{\psi_{J(i,t)}}_{\text{firm eff}} + \underbrace{X_{it}'\beta}_{\text{covariates}} + \varepsilon_{it}$$

- Computational problem: millions of fixed effects. Can't invert $X'X$!
  - Approximate solution method
- Key findings: 90% of industry wage premia attributable to person effects
  - Explanation: industry switching estimates biased by nonrandom sorting of workers to firms within industry (Really??)

Abowd, Creecy, Kramarz (2002)

	Standard Deviation	Correlations								
		lny	$x\beta$	θ	ψ	ε	$x\beta$ (approx.)	θ (approx.)	ψ (approx.)	ε (approx.)
French Data										
Log real annual full-time compensation	0.519	1.000	0.141	0.704	0.201	0.169	0.261	0.840	0.213	0.459
Time-varying characteristics	0.135	0.141	1.000	-0.068	0.023	0.000	0.731	-0.051	0.016	-0.057
Person effect	0.455	0.704	-0.068	1.000	-0.283	0.000	-0.017	0.836	0.021	0.044
Firm effect	0.285	0.201	0.023	-0.283	1.000	0.000	0.036	0.217	0.184	-0.022
Residual	0.206	0.169	0.000	0.000	0.000	1.000	-0.005	0.000	0.048	0.359
Time-varying characteristics (approximate)	0.146	0.261	0.731	-0.017	0.036	-0.005	1.000	0.001	0.019	-0.052
Person effect (approximate)	0.425	0.840	-0.051	0.836	0.217	0.000	0.001	1.000	0.097	0.016
Firm effect (approximate)	0.065	0.213	0.016	0.021	0.184	0.048	0.019	0.097	1.000	0.007
Residual (approximate)	0.238	0.459	-0.057	0.044	-0.022	0.359	-0.052	0.016	0.007	1.000
State of Washington Data										
Log real hourly compensation	0.527	1.000	0.304	0.511	0.518	0.306	0.323	0.585	0.478	0.331
Time-varying characteristics	0.380	0.304	1.000	-0.530	0.143	0.000	0.998	-0.485	0.172	0.000
Person effect	0.476	0.511	-0.530	1.000	-0.025	0.000	-0.512	0.960	0.020	0.000
Firm effect	0.231	0.518	0.143	-0.025	1.000	0.000	0.153	0.155	0.769	0.114
Residual	0.161	0.306	0.000	0.000	0.000	1.000	0.000	0.000	0.000	0.922
Time-varying characteristics (approximate)	0.361	0.323	0.998	-0.512	0.153	0.000	1.000	-0.469	0.181	0.000
Person effect (approximate)	0.470	0.585	-0.485	0.960	0.155	0.000	-0.469	1.000	0.050	0.000
Firm effect (approximate)	0.163	0.478	0.172	0.020	0.769	0.000	0.181	0.050	1.000	0.000
Residual (approximate)	0.175	0.331	0.000	0.000	0.114	0.922	0.000	0.000	0.000	1.000

Notes: The column headers use the symbols from the text while the row headers provide short definitions. All approximations are based on AKM (1999), persons first, formulas.
Sources: Authors' calculations based on the INSEE and State of Washington UI data.

- Approx FEs very weakly correlated with exact FEs in French data $\Rightarrow$ original AKM results invalid!
- Exact results find $\frac{Var(\psi_{J(i,t)})}{Var(y_{it})} \approx 55\%$ in France and 45% in Washington state
- Note: $Cov\left(\psi_{J(i,t)}, \alpha_i\right) < 0$ – negative assortative matching!

Abowd, Lengermann, and McKinney (2003)

Table 6: Summary of Pooled Human Capital Wage Components

Component	Standard Deviation	Correlation with: *ln w*	$x\beta$	θ	α	$\upsilon\eta$	ψ	ε
Log Real Annualized Wage Rate (*ln w*)	0.881	1.000	0.224	0.468	0.451	0.212	0.484	0.402
Time-Varying Personal Characteristics ($x\beta$)	0.691	0.224	1.000	-0.553	-0.575	-0.099	0.095	0.000
Person Effect (θ)	0.835	0.468	-0.553	1.000	0.961	0.275	0.080	0.000
Unobserved Part of Person Effect (α)	0.802	0.451	-0.575	0.961	1.000	0.000	0.045	0.000
Non-time-varying Personal Characteristics ($u\eta$)	0.229	0.212	-0.099	0.275	0.000	1.000	0.101	0.000
Firm Effect (ψ)	0.362	0.484	0.095	0.080	0.045	0.101	1.000	0.000
Residual (ε)	0.354	0.402	0.000	0.000	0.000	0.000	0.000	1.000

Notes: Based on 287,241,891 annual observations from 1986-2000 for 68,329,212 persons and 3,662,974 firms in California, Florida, Illinois, Maryland, Minnesota, North Carolina, and Texas. No single state contributed observations for all years. See Table 1.
Sources: Author's calculations using the LEHD Program data base.

- Use a 100% extract from LEHD of 7 states instead of small subsamples
- Correlation becomes positive! (limited-mobility bias)
- $\frac{Var\left(\psi_{J(i,t)}\right)}{Var(y_{it})} \approx 20\%$ (less inflation due to sampling error)

Critiques of AKM

Not theoretically motivated

- Why same firm effect for different types of workers?
- Why wages monotone in productivity?

Negative assortativeness implausible

Endogenous mobility:

- Selection on match
- Selection on firm shocks

Person and firm effects inconsistent in short panels (Abowd, Creecy, Kramarz, 2002; Andrews, 2008)

- Variances biased upwards (same issue as Krueger-Summers)
- Correlation between FE,PE biased downwards

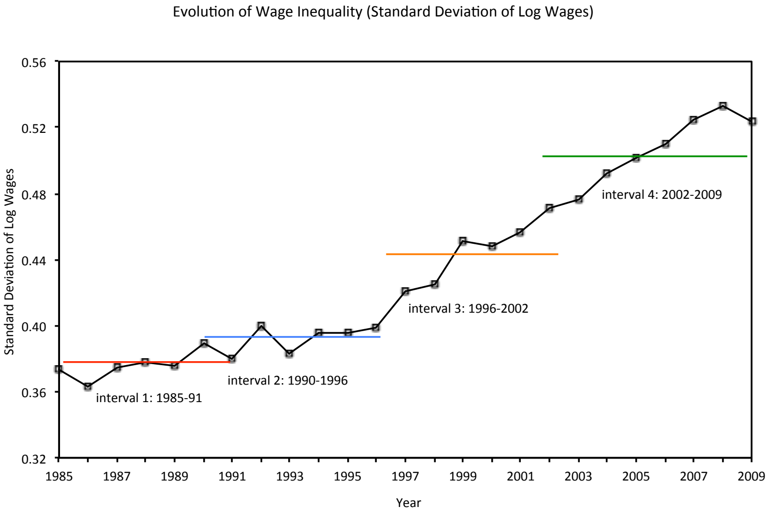
Card, Heining, and Kline (2013)

Study *changes* in German wage structure

- Earlier work by Dustmann, Ludsteck, and Schoenberg (2009) documented an increase in German wage dispersion
- Interpreted within traditional SDI framework – supply / demand / institutions
- Typical view: S+D influence price of skill, I is barrier to price adjustment
- Need SDI-(F) for firms/frictions?

"Rolling"-AKM over 6-7 year intervals

- Each decomposition gives us a "snapshot" of labor market
- Did sorting change? Did importance of firms change?
- Check for endogenous mobility



Interval timing coincides with waves of liberalization of German labor market:

- Labor Law Act for Promotion of Employment: 1996
- Hartz reforms: 2003-2005

Growth in wage inequality primarily between establishments

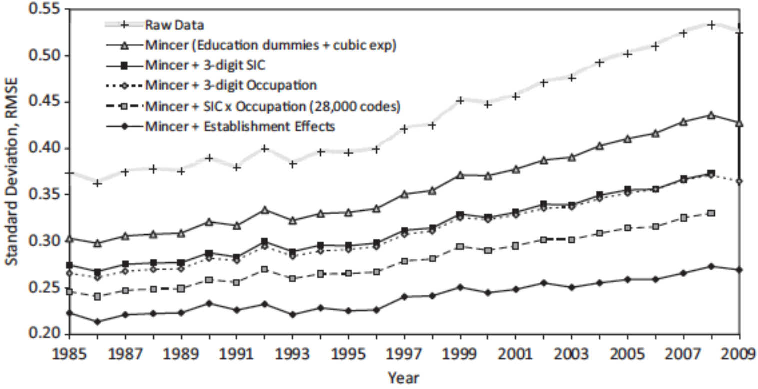


FIGURE IV

Raw and Residual Standard Deviations from Alternative Wage Models

Wage dynamics of job changes

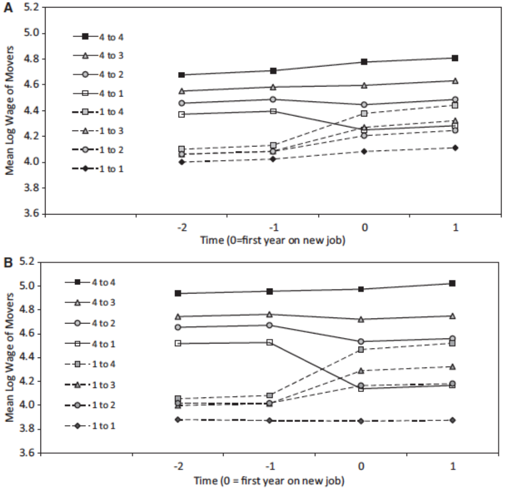


Figure V

Mean Wages of Job Changers Classified by Quartile of Mean Wage of Coworkers at Origin and Destination Establishment (A) 1985–1991, (B) 2002–2009

Table 2: Estimation Results for AKM Model, Fit by Interval

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Dimensions / Summary Stats:				
Number person effects	**16,295,106**	**17,223,290**	**16,384,815**	**15,834,602**
Number establishment effects	**1,221,098**	**1,357,824**	**1,476,705**	**1,504,095**
Sample size (person-year obs)	**84,185,730**	**88,662,398**	**83,699,582**	**90,615,841**
Std. Dev. Log Wages	0.370	0.384	0.432	0.499
Summary of Parameter Estimates:				
Std. dev. of person effects	0.289	0.304	0.327	0.357
Std. dev. of establ. effects	0.159	0.172	0.194	0.230
Std. dev. of Xb	0.121	0.088	0.093	0.084
Correlation of person/establ. effects	0.034	0.097	0.169	0.249
(across person-year obs.)				
RMSE of AKM residual	0.119	0.121	0.130	0.135
(degrees of freedom)	66,669,487	70,081,245	65,838,023	73,277,100
Adjusted R-squared	0.896	0.901	0.909	0.927
Comparison Match Model				
RMSE of Match model	0.103	0.105	0.108	0.112
Adjusted R-squared	0.922	0.925	0.937	0.949
Std. Dev. of Match Effect*	0.060	0.060	0.072	0.075

Table 2: Estimation Results for AKM Model, Fit by Interval

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Dimensions / Summary Stats:				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Sample size (person-year obs)	84,185,730	88,662,398	83,699,582	90,615,841
Std. Dev. Log Wages	**0.370**	**0.384**	**0.432**	**0.499**
Summary of Parameter Estimates:				
Std. dev. of person effects	0.289	0.304	0.327	0.357
Std. dev. of establ. effects	0.159	0.172	0.194	0.230
Std. dev. of Xb	0.121	0.088	0.093	0.084
Correlation of person/establ. effects (across person-year obs.)	0.034	0.097	0.169	0.249
RMSE of AKM residual	0.119	0.121	0.130	0.135
(degrees of freedom)	66,669,487	70,081,245	65,838,023	73,277,100
Adjusted R-squared	0.896	0.901	0.909	0.927
Comparison Match Model				
RMSE of Match model	0.103	0.105	0.108	0.112
Adjusted R-squared	0.922	0.925	0.937	0.949
Std. Dev. of Match Effect*	0.060	0.060	0.072	0.075

Table 2: Estimation Results for AKM Model, Fit by Interval

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Dimensions / Summary Stats:				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Sample size (person-year obs)	84,185,730	88,662,398	83,699,582	90,615,841
Std. Dev. Log Wages	**0.370**	**0.384**	**0.432**	**0.499**
Summary of Parameter Estimates:				
Std. dev. of person effects	**0.289**	**0.304**	**0.327**	**0.357**
Std. dev. of establ. effects	**0.159**	**0.172**	**0.194**	**0.230**
Std. dev. of Xb	0.121	0.088	0.093	0.084
Correlation of person/establ. effects (across person-year obs.)	0.034	0.097	0.169	0.249
RMSE of AKM residual	0.119	0.121	0.130	0.135
(degrees of freedom)	66,669,487	70,081,245	65,838,023	73,277,100
Adjusted R-squared	0.896	0.901	0.909	0.927
Comparison Match Model				
RMSE of Match model	0.103	0.105	0.108	0.112
Adjusted R-squared	0.922	0.925	0.937	0.949
Std. Dev. of Match Effect*	0.060	0.060	0.072	0.075

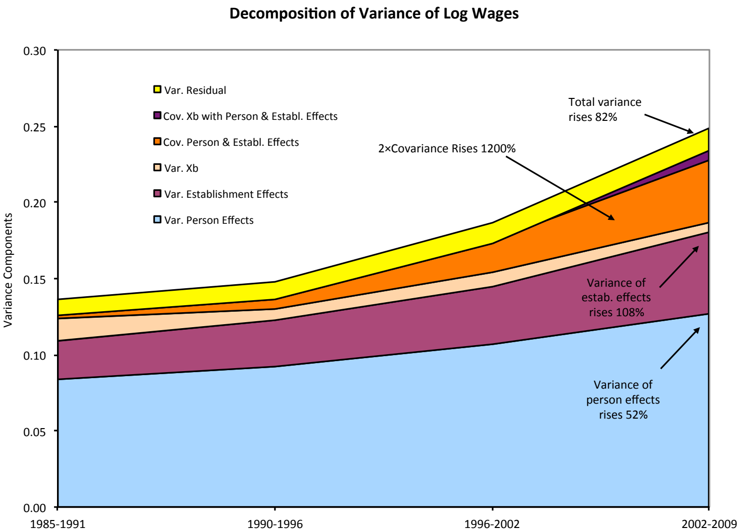
Table 2: Estimation Results for AKM Model, Fit by Interval

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Dimensions / Summary Stats:				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Sample size (person-year obs)	84,185,730	88,662,398	83,699,582	90,615,841
Std. Dev. Log Wages	**0.370**	**0.384**	**0.432**	**0.499**
Summary of Parameter Estimates:				
Std. dev. of person effects	**0.289**	**0.304**	**0.327**	**0.357**
Std. dev. of establ. effects	**0.159**	**0.172**	**0.194**	**0.230**
Std. dev. of Xb	0.121	0.088	0.093	0.084
Correlation of person/establ. effects (across person-year obs.)	**0.034**	**0.097**	**0.169**	**0.249**
RMSE of AKM residual	0.119	0.121	0.130	0.135
(degrees of freedom)	66,669,487	70,081,245	65,838,023	73,277,100
Adjusted R-squared	0.896	0.901	0.909	0.927
Comparison Match Model				
RMSE of Match model	0.103	0.105	0.108	0.112
Adjusted R-squared	0.922	0.925	0.937	0.949
Std. Dev. of Match Effect*	0.060	0.060	0.072	0.075

Table 2: Estimation Results for AKM Model, Fit by Interval

	Interval 1 1985-1991 (1)	Interval 2 1990-1996 (2)	Interval 3 1996-2002 (3)	Interval 4 2002-2009 (4)
Dimensions / Summary Stats:				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Sample size (person-year obs)	84,185,730	88,662,398	83,699,582	90,615,841
Std. Dev. Log Wages	**0.370**	**0.384**	**0.432**	**0.499**
Summary of Parameter Estimates:				
Std. dev. of person effects	**0.289**	**0.304**	**0.327**	**0.357**
Std. dev. of establ. effects	**0.159**	**0.172**	**0.194**	**0.230**
Std. dev. of Xb	0.121	0.088	0.093	0.084
Correlation of person/establ. effects (across person-year obs.)	**0.034**	**0.097**	**0.169**	**0.249**
RMSE of AKM residual	0.119	0.121	0.130	0.135
(degrees of freedom)	66,669,487	70,081,245	65,838,023	73,277,100
Adjusted R-squared	**0.896**	**0.901**	**0.909**	**0.927**
Comparison Match Model				
RMSE of Match model	0.103	0.105	0.108	0.112
Adjusted R-squared	**0.922**	**0.925**	**0.937**	**0.949**
Std. Dev. of Match Effect*	0.060	0.060	0.072	0.075

Growing firm component



Diagnostics

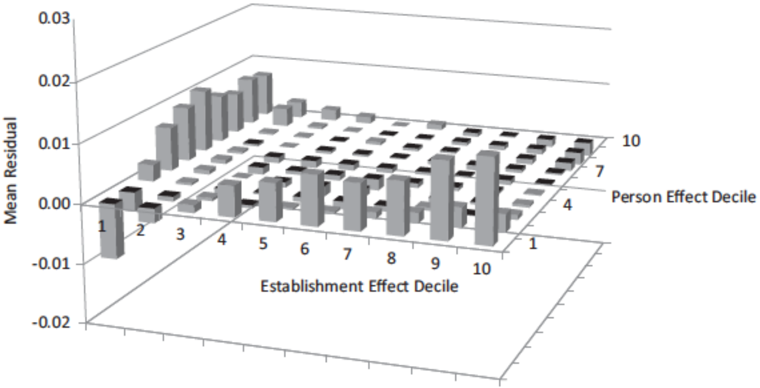
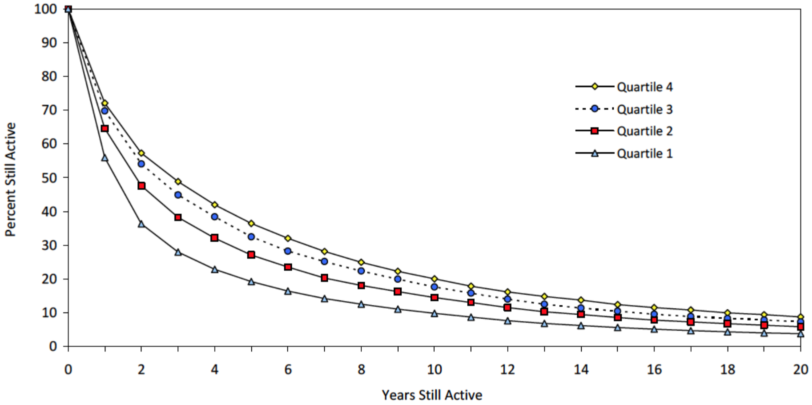


FIGURE VI

Mean Residuals by Person/Establishment Deciles, 2002–2009

High firm effect jobs last longer

Figure 17b: Survivor Functions for Jobs Initiated in 1989
By Quartile of Estimated Establishment Effect



Notes: figure shows fraction of jobs held by full time male workers in IEB that were initiated in 1989 and are still present after number of years indicated by x-axis. Establishments are divided into quartiles based on their estimated establishment effects from an AKM model fit to data from 1985 to 1991. Quartile 1 refers to the lowest quartile of estimated establishment effects.

Change in returns to education largely due to change in estab effect!

TABLE V

DECOMPOSITION OF CHANGES IN RELATIVE WAGES BY EDUCATION LEVEL, 1985–1991 VERSUS 2002–2009

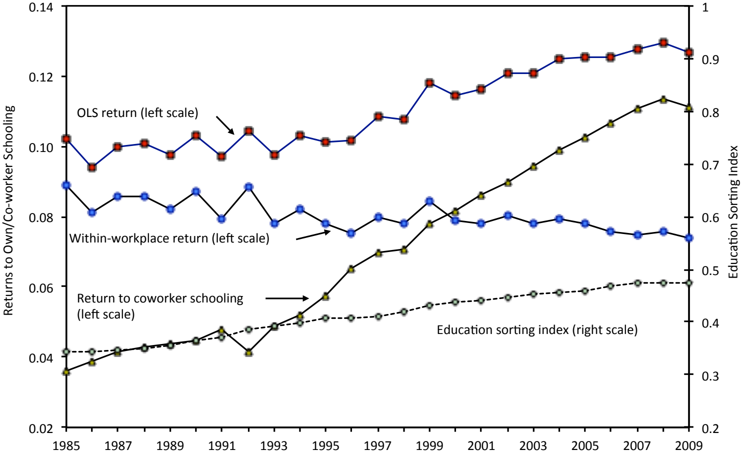
	(1) Change in mean log wage relative to apprentices	(2) Change in mean person effect	(3) Change in mean establishment effect	(4) Remainder
Highest education qualification				
1. Missing/none	–14.6	1.8	–12.2	–4.2
2. Lower secondary school or less (no vocational training)	–10.5	–0.1	–6.3	–4.1
4. Abitur with or without vocational training*	10.1	0.0	2.6	7.5
5. University or more	5.7	1.5	3.9	0.3

Notes. Wage changes are measured between intervals 1 (1985–1991) and 4 (2002–2009). Remainder (column (4)) represents changing relative contribution of Xb component.

**Abitur* refers to Allgemeine Hochschulreife, a certificate of completion of advanced level high school.

Does this reflect changes in the sorting of workers to firms?

Cross-Check: Mundlak (1978) Decomposition of Return to Education



- Sorting index is coefficient from regression of mean schooling at firm ($\bar{S}_{J(i,t)}$) on individual schooling (S_i)
- Mundlak comes from running $w_{it} = \alpha + \beta S_i + \delta \bar{S}_{J(i,t)} + \varepsilon_{it}$

TABLE VI

Contribution of Person and Establishment Effects to Wage Variation across Occupations and Industries

	(1) Interval 1 1985–1991	(2) Interval 2 1990–1996	(3) Interval 3 1996–2002	(4) Interval 4 2002–2009	Change in variance (Int. 1 to Int. 4)* (5)	(6) Share
Panel A: Between occupations (342 three-digit occupations)						
Std. dev. of mean log wages	0.233	0.243	0.263	0.289	0.029	100
Std. dev. of mean person effects	0.186	0.203	0.198	0.207	0.008	28
Std. dev. of mean estbl. effects	0.101	0.104	0.124	0.135	0.008	28
Correlation of mean person effects and establ. effects	0.110	0.171	0.238	0.291	0.012	42
Panel B: Between industries (96 two-digit industries)						
Std. dev. of mean log wages	0.173	0.184	0.203	0.224	0.020	100
Std. dev. of mean person effects	0.103	0.114	0.128	0.140	0.009	44
Std. dev. of mean estbl. effects	0.104	0.110	0.108	0.121	0.004	19
Correlation of mean person effects and establ. effects	0.242	0.301	0.422	0.403	0.008	42

Notes. Decompositions based on estimated AKM models summarized in Table III. Occupation is based on main job in each year; establishments are assigned one industry per interval, using consistently-coded two-digit industry.

*Entry in column (5) represents change in variance or covariance component. Entry in column (6) is the share of the total change in variance explained. Shares do not add to 100% because Xb component and its covariances are omitted.

- Contribution of "pure" person component to variance of (unadjusted) industry wage differences ≈ 35-40%
- Rise in between-group inequality explained by mix of dispersion in person and firm effects
- But biggest contributor is increased correlation (i.e., sorting)

Changes driven by breakdown in bargaining?

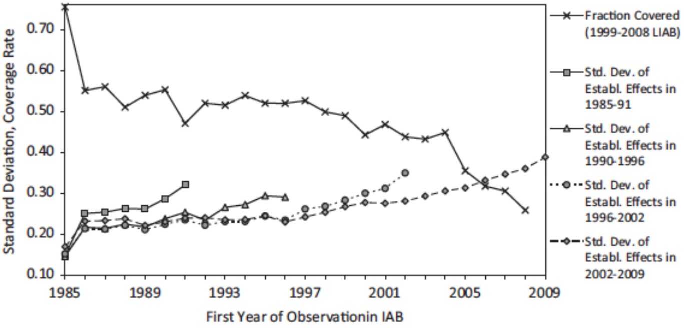


FIGURE IX

Standard Deviation of Establishment Effects and Fraction Covered by Collective Agreements, by Birth Year of Establishment

Note: newer firms more variable regardless of time period!

Low paying firms not covered by collective bargaining

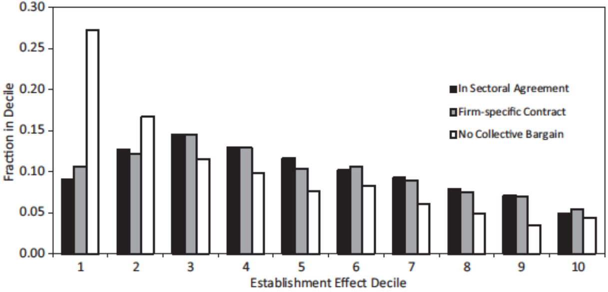


Figure X

Distribution of Establishment Effects by Collective Bargaining Status, Based on Establishment Effects for 1996–2002 and Bargaining Status in 2000 Wave of LIAB

Takeaway

AKM as a tool for studying changes in wage structure

- Decompose traditional wage gaps (education, industry, occupation) into person and firm components
- Maybe endogenous mobility not so bad?

Result: big changes in German labor market

- Firms growing more important both directly (wage effects) and indirectly (sorting)
- Timing lines up with institutional changes
- Major cohort effects in firm inequality

Card, Cardoso, and Kline (2016)

Gender wage gap: women paid less than men

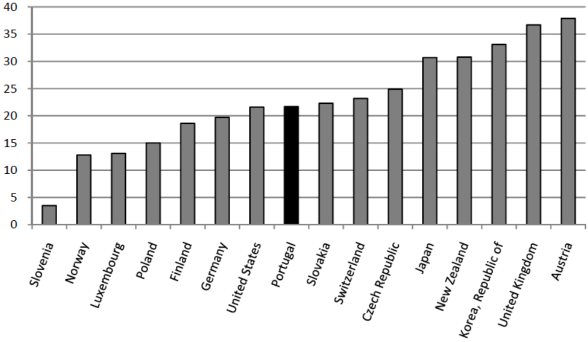
- Traditional explanation: women less skilled
- Alternative hypothesis: nice girls don't ask (Babcock and Laschever, 2009)

Examine outside of the lab by looking at E-E wage data from Portugal merged with firm Value Added measures from BvD

- Q1: do women get the same firm effs as men?
- Q2: do shocks to firm VA get shared equally with male and female employees?

Portugal gender gap similar to US

Figure 2 - Gender pay gap (%)



Source: ILO, ILOSTAT Database, www.ilo.org, accessed July 10, 2013.

But Portuguese women have more schooling than men!

Table 1: Descriptive Statistics for Samples of Employees in QP, 2002-2009

	Full sample age 19-65; exp>1, valid wages/hours/tenure in all years		Workers at dual-connected firms		Full sample with VA data	
	Males	Females	Males	Females	Males	Females
Education (yrs)	8.0	8.8	8.6	9.1	8.1	8.9
Log Real Hrly Wage	1.59	1.41	1.71	1.48	1.57	1.38
(standard dev.)	(0.55)	(0.50)	(0.58)	(0.53)	(0.50)	(0.45)
Monthly Hours	162.6	158.0	162.8	157.1	163.8	159.0
(standard dev.)	(24.7)	(30.1)	(24.0)	(30.5)	(24.5)	(30.8)
Firm Size (#wkrs)	730	858	1,091	1,230	641	1,117
Fraction Female at Firm	0.24	0.70	0.30	0.64	0.24	0.67
Log VA/ Worker					3.08	2.90
Number of person-year obs.	9.07M	7.23M	6.01M	5.01M	3.34M	2.45M
Number of persons	2.12M	1.75M	1.45M	1.25M	1.21M	0.92M
Number of firms	350K	336K	85K	85K	160K	148K

Notes: Overall sample in columns 1-2 includes paid workers age 19-65 with potential experience ≥1. Sampe excludes individuals with inconsistent employment histories. Wages are measured in real (2009=100) Euros per hour. Value added is measured in thousands of real Euros per year. All statistics are calculated across person-year observations. See text for definitions of connected and dual connected sets.

Note also that women work at larger but *less productive* firms. Are they trapped in bad jobs?

Women switch firms about as often as men

Appendix Table B1: Distributions of Number of Jobs Held in Sample Period, by Gender, and Mean Log Wage by Number of Jobs Held

	Distribution of Number of Jobs Held 2002-2009 (Person-year weighted)		Distribution of Number of Jobs Held 2002-2009 (Person-weighted)		Mean Log Wage of Persons, By Number of Jobs Held 2002-2009		
	Males	Females	Males	Females	Males	Females	Male-Female Gap
# Jobs	(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	67.81	70.37	72.50	74.29	1.56	1.38	0.17
2	20.93	20.42	18.71	18.51	1.45	1.31	0.15
3	7.91	6.84	6.39	5.53	1.43	1.29	0.14
4	2.52	1.87	1.85	1.35	1.41	1.28	0.13
5	0.68	0.41	0.46	0.27	1.39	1.27	0.12
6	0.13	0.08	0.08	0.05	1.39	1.26	0.13
7	0.02	0.01	0.01	0.01	1.37	1.22	0.14
8	0.00	0.00	0.00	0.00	1.39	1.48	-0.09
# Obs.	9,070,492	7,226,310	2,119,687	1,747,492	2,119,687	1,747,492	--

Notes: tabulations based on analysis sample of male and female employees in QP data set -- see columns 1 and 2 of Table I. There are 15 males and 7 females with 8 jobs in the sample, accounting for 120 person-year observations for men and 56 person-year observations for women.

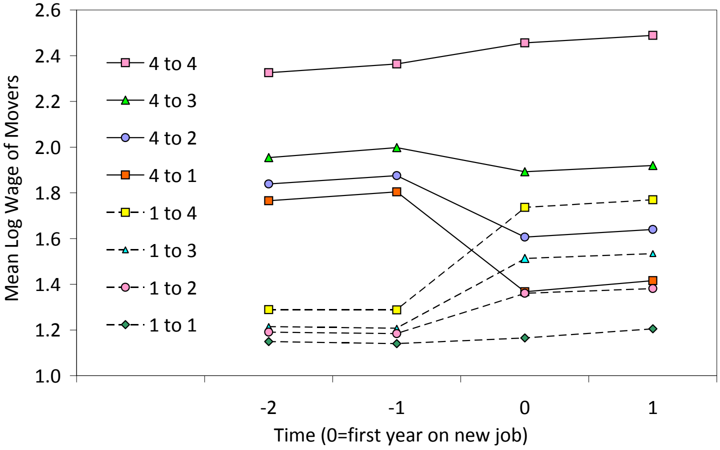
But women more likely to move between low wage firms

Appendix Table B2: Wages of Job Changes for Movers with 2+ Years of Data Before/After Job Change

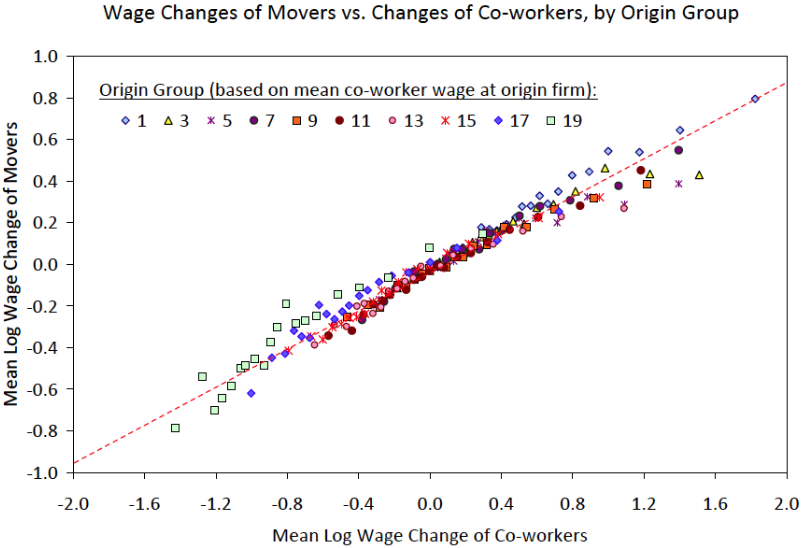
Origin/ destination quartile	Number Changes (1)	Pct. Of Changes (2)	Mean Log Real Wages of Movers: 2 years before (3)	1 year before (4)	1 year after (5)	2 years after (6)	3 Year Change (%) Raw (7)	Adjusted* (8)	(Std Err) (9)
Males									
1 to 1	13,787	43.2	1.14	1.14	1.16	1.20	5.6	0.5	(0.5)
1 to 2	9,139	28.7	1.19	1.18	1.35	1.37	17.6	11.6	(0.6)
1 to 3	6,283	19.7	1.20	1.19	1.48	1.51	30.6	23.9	(0.7)
1 to 4	2,682	8.4	1.28	1.27	1.71	1.75	47.3	39.0	(1.2)
2 to 1	7,293	21.2	1.34	1.35	1.22	1.27	-6.5	-12.0	(0.6)
2 to 2	12,326	35.8	1.37	1.38	1.40	1.42	5.0	-0.8	(0.6)
2 to 3	10,356	30.0	1.41	1.42	1.54	1.57	15.9	9.3	(0.5)
2 to 4	4,496	13.0	1.49	1.49	1.81	1.84	35.3	27.0	(0.9)
3 to 1	4,356	11.9	1.49	1.52	1.24	1.30	-19.4	-25.6	(0.7)
3 to 2	8,835	24.2	1.54	1.55	1.45	1.48	-5.8	-12.2	(0.6)
3 to 3	15,107	41.3	1.61	1.63	1.65	1.67	6.4	-0.3	(0.5)
3 to 4	8,246	22.6	1.73	1.75	1.94	1.97	24.7	16.0	(0.7)
4 to 1	1,634	5.4	1.79	1.83	1.39	1.43	-36.2	-43.3	(1.6)
4 to 2	3,245	10.7	1.82	1.86	1.58	1.61	-20.9	-28.1	(1.2)
4 to 3	6,589	21.7	1.93	1.97	1.85	1.88	-5.2	-13.1	(0.9)
4 to 4	18,830	62.1	2.29	2.32	2.41	2.45	15.9	6.1	(0.9)
Females									
1 to 1	24,130	60.9	1.05	1.04	1.05	1.08	2.9	-0.6	(0.4)
1 to 2	9,094	23.0	1.10	1.10	1.21	1.23	13.2	8.4	(0.5)
1 to 3	4,490	11.3	1.13	1.14	1.35	1.37	23.6	17.6	(0.6)
1 to 4	1,888	4.8	1.25	1.26	1.59	1.62	37.0	29.6	(1.2)
2 to 1	6,705	29.8	1.20	1.22	1.12	1.16	-4.5	-9.1	(0.5)
2 to 2	7,711	34.3	1.26	1.28	1.28	1.31	4.2	-1.2	(0.5)
2 to 3	5,495	24.5	1.33	1.35	1.44	1.46	12.6	6.4	(0.8)
2 to 4	2,562	11.4	1.44	1.45	1.69	1.73	29.0	20.7	(0.9)
3 to 1	3,283	16.7	1.38	1.40	1.15	1.20	-17.4	-23.0	(1.3)
3 to 2	4,762	24.2	1.42	1.45	1.34	1.37	-4.5	-10.9	(1.1)
3 to 3	7,245	36.8	1.51	1.53	1.54	1.56	5.3	-1.2	(0.7)
3 to 4	4,381	22.3	1.64	1.66	1.81	1.86	22.0	13.4	(0.9)
4 to 1	1,014	6.2	1.60	1.64	1.32	1.36	-24.6	-31.3	(2.8)
4 to 2	1,516	9.2	1.72	1.76	1.54	1.58	-13.7	-21.2	(1.3)
4 to 3	2,844	17.3	1.82	1.86	1.76	1.81	-1.3	-9.3	(0.9)
4 to 4	11,064	67.3	2.14	2.18	2.27	2.31	16.1	7.0	(0.8)

Notes: entries are mean log real wages for job changers to/from mixed-gender firms with at least 2 years of wages at the old job and the new job. Origin/destination quartiles are based on mean wages of coworkers in year before (origin) or year after (destination) job move.

Figure 2a: Mean Wages of Male Job Changers By O/D Co-worker Group

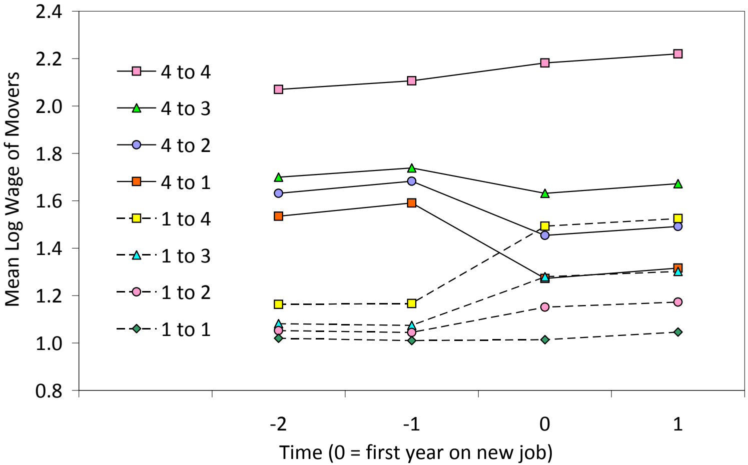


A closer look

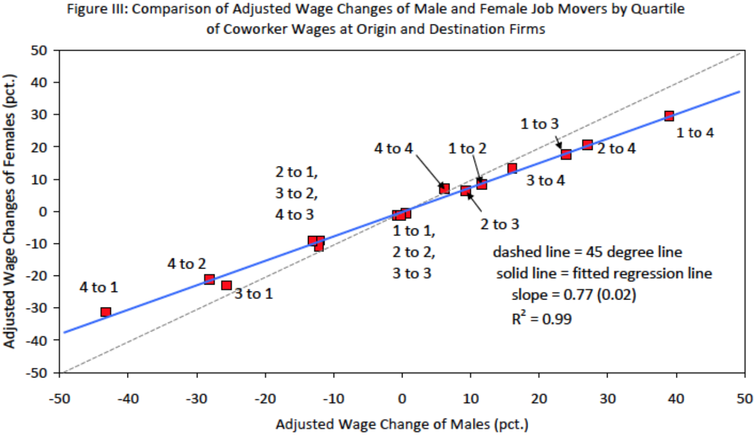


$$E\left[\Delta w | \Delta w^{\text{coworker}}\right] \approx 0.4 \Delta w^{\text{coworker}}$$

Figure 2b: Mean Wages of Female Job Changers by O/D Coworker Group



Women's wages less sensitive to firm rank than men's



Notes: points represent regression adjusted mean log wage changes of male and female job movers in different origin/ destination quartiles of mean coworker wages. For example "4 to 1" point shows mean wage changes for men and women who move from 4th quartile of coworker wages to 1st quartile. Fitted line is estimated by OLS to 16 points in the Figure.

Gender-specific AKMs

Table 3: Summary of Estimated Models for Male and Female Workers

	Males	Females	German Men
Summary of Parameter Estimates: AKM Model			
Std. dev. of pers. effects (person-yr obs.)	0.420	0.400	0.357
Std. dev. of firm effects (person-yr obs.)	0.247	0.213	0.230
Std. dev. of Xb (across person-yr obs.)	0.069	0.059	0.084
Correlation of person/firm effects	0.167	0.152	0.249
Adjusted R-squared	0.934	0.940	0.927
Correlation male / female firm effects	**0.590**		
Comparison job-match effects model:			
Adjusted R-squared	0.946	0.951	0.949
Std. deviation match effect in AKM model	0.062	0.054	0.075
Share of variance of log wages due to:			
person effects	57.6	61.0	51.2
firm effects	19.9	17.2	21.2
covariance of person/firm effects	11.4	9.9	16.4
Xb and associated covariances	6.2	7.5	5.2
residual	4.9	4.4	5.9

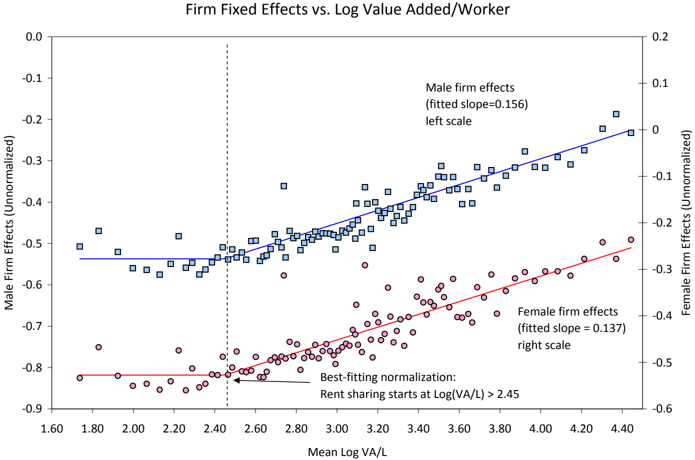
No evidence of compensating diff for hours

Appendix Table B7: Relationship Between Estimated Firm Effects and Mean Hours of Workers of Same Gender

	Models for Males				Models for Females			
	No Industry Controls		Industry Controls		No Industry Controls		Industry Controls	
	OLS	IV	OLS	IV	OLS	IV	OLS	IV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Using Regular Contractual Hours								
Log Mean Hours of Workers at Firm (Same Gender)	-0.22	-0.13	-0.11	0.01	-0.06	-0.24	0.02	-0.07
	(0.04)	(0.05)	(0.03)	(0.05)	(0.03)	(0.05)	(0.02)	(0.04)
First Stage Coeff.	--	0.52	--	0.43	--	0.68	--	0.63
		(0.00)		(0.01)		(0.00)		(0.00)
B. Using Total Hours								
Log Mean Hours of Workers at Firm (Same Gender)	-0.16	-0.12	-0.06	0.02	-0.05	-0.13	0.03	0.03
	(0.03)	(0.05)	(0.03)	(0.05)	(0.03)	(0.05)	(0.02)	(0.04)
First Stage Coeff.	--	0.54	--	0.45	--	0.65	--	0.60
		(0.00)		(0.01)		(0.00)		(0.00)

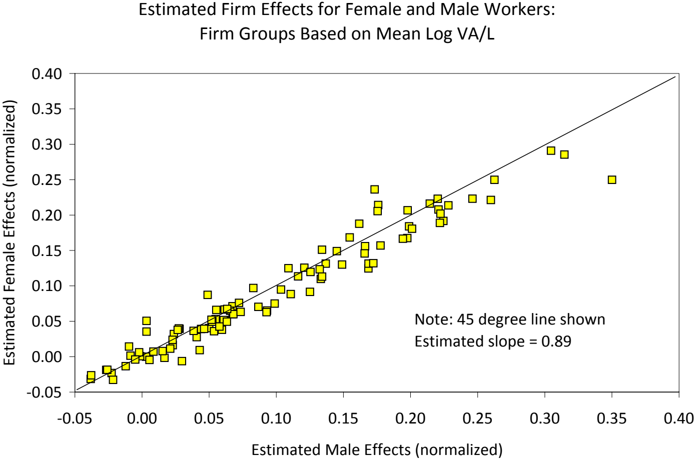
Notes: Dependent variable in columns 1-4 is estimated firm-specific wage premium for male employees at a firm. Dependent variable in columns 5-8 is estimated firm-specific wage premium for female employees. Entries represent coefficients of log mean hours of the gender group at the firm. Hours measure in Panel A is regular contractual hours. Hours measure in Panel B is total hours. Models in columns 3-4 and 7-8 include dummies for 20 major industries. All specifications include a constant. Models in even-numbered columns are estimated by IV, using the log mean hours of workers at the same firm in the other gender group as an instrument. Estimated first stage coefficients are reported in second row of the table. All models are fit to micro data for workers in the dual-connected set (n=11,025,257), with standard errors (in parentheses) clustered by firm (n=84,720 firms).

"Hockey stick" relationship of FEs with productivity



- Normalize gender specific FEs=0 below kink to compare levels (below kink is "competitive frontier")
- Female FEs have lower VA elasticity. Ratio $= 1.37/1.56 \approx 0.9$

Grouping estimate of relative rent sharing = 0.89



Note: implicitly using VA/L as instrument for male FEs here

Oaxaca review

Give women male firm effects

$$E[\psi^M_{J(i,t)}|G(i)=M] - E[\psi^F_{J(i,t)}|G(i)=F] = \underbrace{E[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|G(i)=F]}_{\text{Bargaining}}$$
$$+ \underbrace{E[\psi^M_{J(i,t)}|G(i)=M] - E[\psi^M_{J(i,t)}|G(i)=F]}_{\text{Sorting}}$$

Assign men to same firms as women

Or Equivalently:

Give men female firm effects

$$E[\psi^M_{J(i,t)}|G(i)=M] - E[\psi^F_{J(i,t)}|G(i)=F] = \underbrace{E[\psi^M_{J(i,t)} - \psi^F_{J(i,t)}|G(i)=M]}_{\text{Bargaining}}$$
$$+ \underbrace{E[\psi^F_{J(i,t)}|G(i)=M] - E[\psi^F_{J(i,t)}|G(i)=F]}_{\text{Sorting}}$$

Assign women to same firms as men

Table 4a. Contribution of Firm-based Wage Components to Male-Female Wage Gap

	Gender Group: Males (1)	Gender Group: Females (2)	Difference: Males−Females (percent of overall gap) (3)
1. Mean log wage of group	1.715	1.481	0.234 (100.0)
<u>Means of Estimated Firm Effects:</u>			
2. Firm Effect for Males	0.148	0.114	0.035 (14.9)
3. Firm Effect for Females	0.145	0.099	0.047 (19.9)
4. Within-group Difference in Mean Effects for Males and Females (percent of overall gap)	0.003 (1.2)	0.015 (6.3)	
5. Mean Male Firm Effect for Men minus Mean Female Firm Effect for Women (Total contribution of Firm-based Wage Components)			0.049 (21.2)
6. Sample sizes	6,012,521	5,012,736	

Estimates of sorting effect (using male or female firm effects)

Estimates of differential bargaining power effect (using male or female firm distributions)

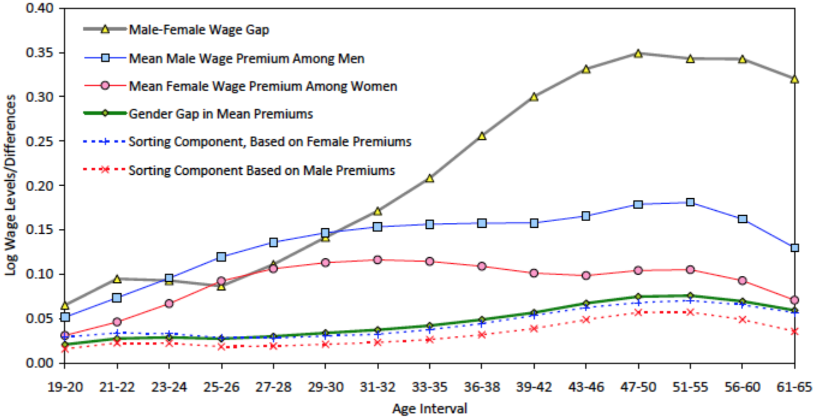
Total contribution of firm components to gender gap

Contribution of Firm-Level Pay Components to Gender Wage Gap

	Gender Wage Gap	Total Contribution of Firm Components	Decompositions			
			Sorting		Bargaining	
			Using M Effects	Using F Effects	Using M Distribution	Using F Distribution
All	-0.234	0.049	0.035	0.047	0.003	0.015
		(21.2)	(14.9)	(19.9)	(1.2)	(6.3)
<u>By Age Group:</u>						
Up to age 30	-0.099	0.028	0.019	0.029	-0.001	0.009
		(28.2)	(18.9)	(29.3)	(-1.2)	(9.3)
Ages 31-40	-0.228	0.045	0.029	0.040	0.004	0.016
		(19.7)	(12.6)	(17.8)	(1.9)	(7.0)
Over Age 40	-0.336	0.069	0.050	0.064	0.005	0.019
		(20.6)	(15.0)	(19.1)	(1.5)	(5.6)
<u>By Education Group:</u>						
< High School	-0.286	0.059	0.045	0.061	-0.002	0.015
		(20.8)	(15.6)	(21.4)	(-0.6)	(5.2)
High School	-0.262	0.061	0.051	0.051	0.010	0.010
		(23.3)	(19.6)	(19.5)	(3.8)	(3.7)
University	-0.291	0.047	0.025	0.029	0.018	0.022
		(16.1)	(8.7)	(9.9)	(6.2)	(7.4)

Notes: see text. Counterfactuals based on estimated two-way fixed effects models described in Table 3.

Figure VI: Evolution of Gender Wage Gap and Its Components Over the Lifecycle



Notes: figure shows unadjusted male-female wage gap, means of firm-specific wage premiums earned by men and women, and the difference in mean premiums, which is the total contribution of firm-specific wage components to the gender wage gap. Dashed lines show the effect of differential sorting of males and females to specific firms, evaluated using male and female firm-specific wage premiums.

Sorting effect sets in gradually over 20s

Quick recap

- Male/Female firm effects highly correlated
- But women seem to only get 90% of the firm effect of men
- 5 log point gap in firm effs between genders
- Oaxaca decomp finds most of firm eff contribution occurs due to women being at different firms than men
- But large unexplained component for higher skilled women

Next: validate with rent sharing estimates for job-stayers

Still get 0.9 ratio when looking at shocks to firm stayers!

Table VI: Effects of Changes in Measured Surplus per Worker on the Change in Wages of Stayers

		Estimated Rent Sharing Coefficients:		Ratio:
	Number of Firms	Male Stayers	Female Stayers	Column (3) / Column (2)
	(1)	(2)	(3)	(4)
Surplus Measure and Sample:				
1. Excess Log Value Added per Worker (Winsorized at +/- 0.50). Sample = Stayers at Firms with Value Added Data, 2006-9	33,104	0.049 (0.007)	0.045 (0.008)	0.911 (0.086)
2. Excess Log Value Added per Worker (Not Winsorized). Sample = Stayers at Firms with Value Added Data, 2006-9	33,104	0.035 (0.006)	0.031 (0.006)	0.894 (0.091)
3. Excess Log Sales per Worker (Winsorized at +/- 0.50). Sample = Stayers at Firms with Sales Data, 2005-8	44,266	0.021 (0.006)	0.018 (0.005)	0.876 (0.182)

Notes: Dependent variables are average change in wages of male or female workers at a firm (regression-adjusted for quadratic in age). Table entries are coefficients of the measured change in surplus per worker, as defined in row heading. Ratios in column 4 are estimated by instrumental variables, treating average change in wages of female stayers as the dependent variable, average change in wages of male stayers as the endogenous explanatory variable, and the change in surplus measure as the instrument. Standard errors, clustered by firm, in parentheses.

Summary

Results consistent with women being less aggressive negotiators (explains $\approx$ 20% of gender wage gap)

- Wage ladder is "taller" for men than women – women only get 90% of male return to moving up a rung on the ladder
- And women seem to have more trouble climbing the ladder than men – their moves aren't as directed up the ladder
- Even among women who stay at the same firm – a shock yields a larger effect on male than female wages.

Do other classic wage gaps (age, race) have firm component?

- IQ (Fredriksson, Hensvik, Skans, 2015)
- Elite education (Huneeus et al., 2015)
- Race (Gerard, Lagos, Severnini, and Card, 2018)

Outsourcing

Workers don't like inequality

Solution: break certain occupations off into a new firm

Weil (2014): the "fissured" workplace

> *Wage discrimination is rarely seen in large firms despite the benefits it could confer. As long as workers are under one roof, the problems presented by horizontal and vertical equity remain. But what if the large employer could wage discriminate by changing the boundary of the firm?*

Goldschmidt and Schmeider (2017)

Study "on site" outsourcing in Germany using administrative records from IAB

Focus on Food Cleaning, Security, and Logistics (FCSL) as occupations most likely to be outsourced

Identify outsourcing events as when a large group of workers leave a "mother" establishment to start a new "daughter" establishment

- Flow of 10+ employees
- Daughter must be a FCSL firm offering business services

FCSL jobs gradually being outsourced

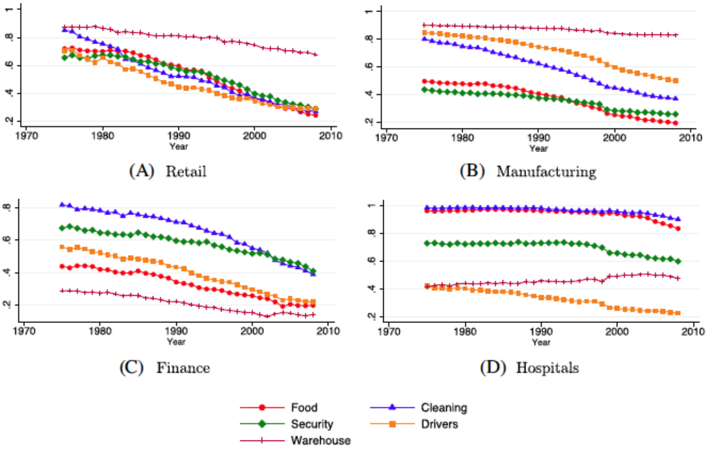
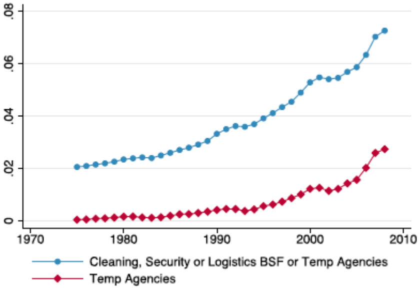


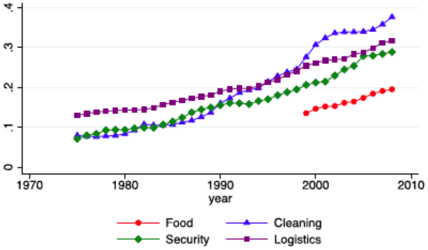
FIGURE I

Share of Firms with any Food/Cleaning/Security/Logistics workers, by Industry

Temp agencies and FCSL firms on the rise



(A) Worker in all Occupations

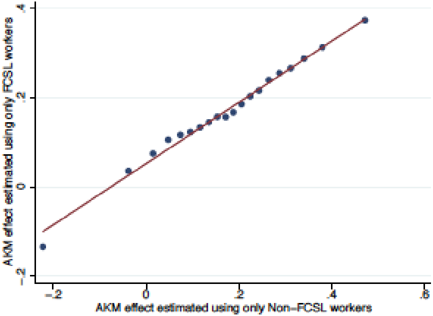


(B) Workers in Food / Cleaning / Security / Logistics Occupations

Figure II

AKM FEs for FCSL and non-FCSL highly correlated

Figure A-8: Comparing Estimated Wage Premia (AKM Effects) based on FCSL and Non-FCSL workers



Notes: The figure shows a binned scatter plot of AKM effects estimated using food, cleaning, security and logistics (FCSL) workers and non-FCSL workers. Both sets of AKM effects are normalized to have a mean of zero in the overall establishment distribution. Each dot corresponds to 1/20th of the observations. Sample is restricted to all German establishments with at least 50 employees.

Being outsourced lowers wages

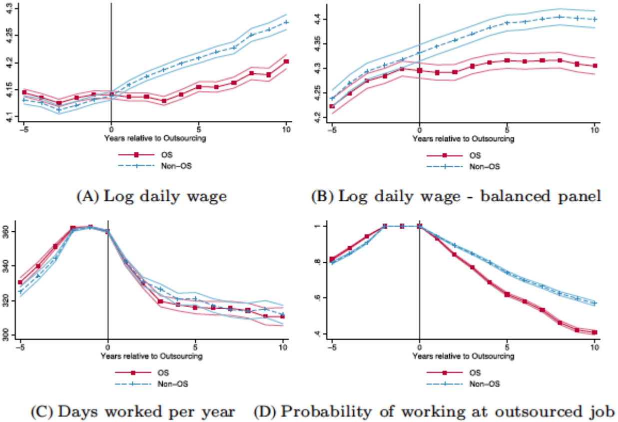


FIGURE IV

Employment Outcomes of Outsourced and Nonoutsourced Workers before and after On-site Outsourcing

Wage losses on order of 10%

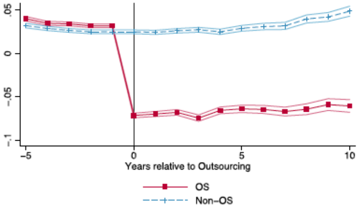
TABLE II

THE EFFECTS OF OUTSOURCING ON LOG DAILY WAGES

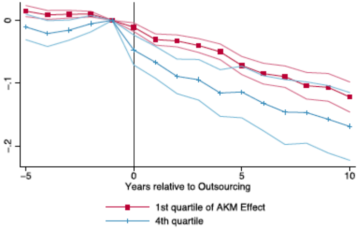
	All FCSL OS events & workers	Food	Cleaning	Security	Logistics	Temp	OS to new estab.	OS to existing estab.
nel A: Effect of on-site outsourcing on workers								
Post-OS short-run	−0.056***	−0.048***	−0.11***	−0.069***	−0.039***	−0.15***	−0.043***	−0.074***
	(0.0048)	(0.0056)	(0.013)	(0.016)	(0.0058)	(0.019)	(0.0053)	(0.0083)
Post-OS long-run	−0.085***	−0.087***	−0.12***	−0.10***	−0.066***	−0.16***	−0.080***	−0.092***
	(0.0077)	(0.010)	(0.019)	(0.021)	(0.010)	(0.019)	(0.0095)	(0.013)
Observations	517,662	158,971	73,064	83,574	202,053	97,538	305,315	212,347
Avg outcome var at $t = -1$	4.14	4.02	3.95	4	4.37	4.37	4.11	4.19
nel B: Effect of on-site outsourcing on jobs								
Post-OS short-run	−0.054***	−0.045***	−0.10***	−0.072***	−0.035***	−0.15***	−0.041***	−0.073***
	(0.0050)	(0.0049)	(0.013)	(0.019)	(0.0057)	(0.016)	(0.0056)	(0.0087)
Post-OS long-run	−0.097***	−0.11***	−0.12***	−0.14***	−0.059***	−0.16***	−0.090***	−0.11***
	(0.0079)	(0.0093)	(0.018)	(0.024)	(0.011)	(0.023)	(0.0099)	(0.013)
Observations	429,949	134,005	61,276	69,976	164,692	72,854	259,434	170,515
Avg outcome var at $t = -1$	4.14	4.02	3.95	4	4.37	4.37	4.11	4.19
nel C: Effects of working for business service firm (Dube and Kaplan 2010 measure)								
Working for business service firm	−0.090***	−0.036***	−0.17***	−0.12***	−0.028***	−0.26***		
	(0.00064)	(0.0030)	(0.0015)	(0.0027)	(0.00064)	(0.00075)		
Observations	36,234,249	1,455,432	10,703,132	3,373,983	20,701,702	13,084,766		
OS workers	1,529,268	45,950	723,294	204,031	576,039	629,278		
Mean outcome for OS workers	3.83	3.79	3.43	3.95	4.21	3.93		

Notes. Standard errors in Panel A and B are clustered at the level of the outsourcing establishment, in Panel C at the worker level. Panels A and B use matched sample of OS l non-OS workers. Panel B includes only workers who are at the same establishment as in time $t = -1$ in all years prior to outsourcing, and in the same establishment as in time 0 in all years after outsourcing. Time periods are five years pre-OS; four years short-run; six years long-run. First column, for all outsourcing types, does not include workers sourced to temp firms. All regressions include individual fixed effects and year dummies, and exclude East Germany before 1997. Panel C, first column includes only workers in d, cleaning, security, or logistics occupations; columns (2)–(5) include only workers in the occupation indicated by the column heading. For food workers, the independent variable a value of 1 if the worker is employed by a firm that provides food services to other companies (defined analogously for other occupations). The sixth column is restricted to ividuals in any occupation who have worked in the same occupation at both a temp agency and in another industry; the independent variable has a value of 1 if the worker employed by a temp agency. All regressions in Panel C control for individual fixed effects, year indicator variables, age, age squared, and age cubed interacted with education mmies; samples are restricted to workers age 25–55, working at establishments with at least 50 workers, and excluding East Germany before 1997. Food workers employed at taurants and hotels are omitted.$^*p < .1$, $^{**}p < .05$, $^{***}p < .01$.

Wage losses entirely explained by drop in AKM FE



(A) Evolution of AKM Effects Before and After Outsourcing



(B) Wage Losses by AKM Effect of Outsourcing Establishment (1st / bottom vs. 4th / top Quartile)

Figure VI

On-site Outsourcing and Establishment (AKM) Effects

Panel A shows the average estimated establishment (AKM) effect of the establishments where the workers in the outsourced and control groups are working before ($t = -1$) and after ($t = 0$) the outsourcing event. The AKM effect is estimated from a wage regression including a full set of worker and establishment fixed effects using the universe of wage records for full-time male workers in Germany. Panel B shows regression estimates of the effects of being outsourced on log wages before and after the outsourcing event separately for workers who are outsourced

Big firms and high wage firms outsource

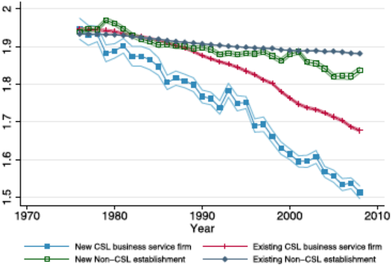
TABLE III

THE EFFECT OF PROXIES FOR WAGE PREMIA ON THE PROBABILITY OF OUTSOURCING

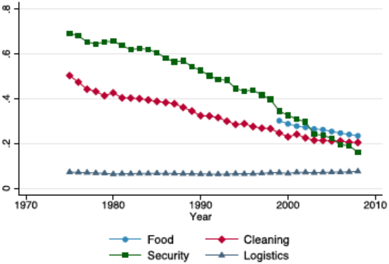
	All establishments				Estab. Panel	
	(1)	(2)	(3)	(4)	(5)	(6)
Log estab size	0.0084***					
	(0.00016)					
Log avg estab wage		0.00044				
		(0.00032)				
AKM effect			0.0046***			
			(0.00057)			
Wage premium to FSCL workers over BSF firms				0.0015***		
				(0.00026)		
Collective agreement					0.0091***	
					(0.0013)	
Pay wages above standard						0.0029**
						(0.0014)
Observations	2,086,507	2,086,505	1,892,408	1,769,077	68,577	68,595
Mean of dep var	0.012	0.012	0.011	0.014	0.02	0.02
Mean of indep var	4.788	4.285	0.003	1.162	0.81	0.34

Notes. Standard errors, in parentheses, are clustered at the establishment level. All regressions exclude East Germany before 1997 and establishments with fewer than 50 workers. Columns (5)–(6) include only establishments included in the IAB Establishment Panel Survey. All regressions control for state dummies, year dummies, and three-digit industry fixed effects. Dependent variable = 1 if the establishment was involved in either a general outsourcing event or an on-site outsourcing event in the following year, and 0 otherwise. "Collective agreement" = 1 if the establishment responded that they were bound by a collective agreement. "Pay wages above standard" = 1 if the establishment responded that they pay salaries and wages above the collectively agreed scale. "Wage premium to FSCL workers over BSF firms" is the ratio of the average wage paid to food, security, cleaning, and logistics workers at the establishment to the average wage paid to food, security, cleaning, and logistics workers employed by business services firms (BSF) or temp agencies in the same county and year.*$p < .1$, **$p < .05$, ***$p < .01$.

Outsourcing a mediating factor for firm cohort effects?



(A) AKM Effects of New and Existing Establishments by Year



(B) Market Concentration of Business Service Firms by Year

Outsourcing explains 7-9% of growth in log wage variance

TABLE IV

THE EVOLUTION OF THE WEST GERMAN WAGE STRUCTURE FROM 1985 TO 2008 AND THE ROLE OF OUTSOURCING

	Wage structure 1985	Wage structure 2008	Change from 1985 to 2008	Percent of change explained
Panel A: Observed				
Total variance of log daily wages	0.132	0.205	0.073	
Variance of estab effects	0.0289	0.0547	0.0258	
2 × cov(person, estab effect)	−0.0050	0.0426	0.0475	
85-15 log wage gap	0.655	0.934	0.279	
85-50 log wage gap	0.385	0.512	0.127	
50-15 log wage gap	0.270	0.422	0.152	
Panel B: Counterfactual I: DFL reweighting of CSL workers				
Total variance of log daily wages	0.132	0.198	0.067	8.9
Variance of estab effects	0.0289	0.0525	0.0236	8.4
2 × cov(person, estab effect)	−0.0050	0.0381	0.0431	9.4
85-15 log wage gap	0.655	0.916	0.260	6.7
85-50 log wage gap	0.385	0.503	0.118	7.1
50-15 log wage gap	0.270	0.412	0.142	6.4
Panel C: Counterfactual II: adjusting daily wage and AKM effect of additional outsourced workers				
Total variance of log daily wages	0.132	0.200	0.068	7.1
Variance of estab effects	0.0289	0.0518	0.0229	11.2
2 × cov(person, estab effect)	−0.0050	0.0408	0.0457	3.8
85-15 log wage gap	0.655	0.925	0.270	3.3
85-50 log wage gap	0.385	0.510	0.125	1.6
50-15 log wage gap	0.270	0.415	0.144	4.7
Percent working in CLS occupations	0.127	0.138	0.011	
Percent outsourced	0.039	0.099	0.060	

Notes. Sample are all full-time male workers in West Germany, excluding workers in food occupations or food industries. Panel A shows the observed wage structure in 1985 and 2008 as well as the estimated components due to the variance of establishment effects and the covariance of establishment with person effects. 85–15 log wage gap refers to the difference between the 85th and 15th percentiles of log daily wages. Panel B shows the counterfactual where workers in cleaning, security, and logistics (CSL) occupations in 2008 are reweighted to keep them at the same percentiles of the AKM distribution as in 1985 using DFL reweighting (see text). Panel C shows the counterfactual where a random fraction of workers in CSL business service firms and temp agencies are "insourced" in 2008 by adding 10 log points to their log wage and establishment effect. The fraction to be insourced is picked so that the fraction of outsourced workers remains at the 1985 level.

Thoughts

Boundaries of the firm are changing

- Easier to pay workers less by segregating them in new establishment
- Wage losses of "fissuring" largely explained by AKM FE
- Validation of causal interpretation

Related literature echoing Gibbons-Katz (1992) uses AKM FE to explain wage effects of mass layoffs

- In Germany AKM FEs explain nearly all of wage loss (Schmieder, Von Wachter, Heining, 2018)
- In Washington state, FEs explain $\sim$ 17% of wage loss (Lachowska, Mas, Woodbury, 2020).
- Important differences in structure of job losses between countries? (Bertheau et al, 2021)

Econometrics of AKM

$$Y_{it} = \alpha_i + \psi_{j(i,t)} + X_{it}'\xi + \varepsilon_{it}$$

where $j(i,t) \in \{1, ..., J\}$ gives identity of current employer.

Matrix representation:

$$Y = D\alpha + F\psi + X\xi + \varepsilon$$

- Isomorphic to standard panel model but with J treatments.
- Treat $Z = (D, F, X)$ as fixed (i.e. all expectations conditional on Z)

Identification:

- Exogeneity: $\mathbb{E}[\varepsilon] = 0$ (plausible?)
- Rank condition: need at least one restriction on the $\{\psi_j\}_{j=1}^J$ within each "connected set" of firms

Variance decomposition

Target parameter: size weighted variance of firm effects

$$\theta_\psi = \sum_{j=1}^{J} s_j \left(\psi_j - \bar{\psi}\right)^2,$$

where s_j is firm j's employment share and $\bar{\psi} = \sum_{j=1}^{J} s_j \psi_j$.

Customary to use OLS estimates $\hat{\psi}$ to compute "plug-in" estimates of variance components, e.g.:

$$\begin{aligned}\hat{\theta}_\psi &= \sum_{j=1}^{J} s_j \left(\hat{\psi}_j - \hat{\bar{\psi}}\right)^2 \\ &= \sum_{j=1}^{J} s_j \left(\hat{\psi}_j\right)^2 - \left(\hat{\bar{\psi}}\right)^2\end{aligned}$$

Bias in the square

OLS is unbiased

$$\mathbb{E}\left[\hat{\psi}_j\right] = \psi_j$$

But the square of an unbiased estimator is upward biased

$$\begin{aligned}
\mathbb{E}\left[\left(\hat{\psi}_j\right)^2\right] &= \mathbb{E}\left[\left(\hat{\psi}_j - \psi_j + \psi_j\right)^2\right] \\
&= \mathbb{E}\left[\left(\hat{\psi}_j - \psi_j\right)^2\right] + 2\mathbb{E}\left[\hat{\psi}_j - \psi_j\right]\psi_j + \psi_j^2 \\
&= \psi_j^2 + \underbrace{\mathbb{V}\left[\hat{\psi}_j\right]}_{\text{bias}}
\end{aligned}$$

Bias of plugin

By same argument plug-in estimator is biased

$$
\begin{aligned}
\mathbb{E}\left[\hat{\theta}_\psi\right] &= \sum_{j=1}^{J} s_j \mathbb{E}\left[\left(\hat{\psi}_j\right)^2\right] - \mathbb{E}\left[\left(\hat{\bar{\psi}}\right)^2\right] \\
&= \sum_{j=1}^{J} s_j \left\{\psi_j^2 + \mathbb{V}\left[\hat{\psi}_j\right]\right\} - (\bar{\psi})^2 - \mathbb{V}\left[\hat{\bar{\psi}}\right] \\
&= \theta_\psi + \underbrace{\sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right] - \mathbb{V}\left[\hat{\bar{\psi}}\right]}_{\text{bias}}
\end{aligned}
$$

$\mathbb{V}\left[\hat{\bar{\psi}}\right]$ term typically negligible when J is large..

Correcting the bias

Bias is weighted average of squared standard errors on firm effects:

$$\mathbb{E}\left[\hat{\theta}_\psi - \theta_\psi\right] \approx \sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right]$$

Correcting the bias

Bias is weighted average of squared standard errors on firm effects:

$$\mathbb{E}\left[\hat{\theta}_\psi - \theta_\psi\right] \approx \sum_{j=1}^{J} s_j \mathbb{V}\left[\hat{\psi}_j\right]$$

Can't we just do Krueger-Summers style correction based on conventional het-consistent ("robust") standard errors $\hat{\mathbb{V}}_{HC}\left[\hat{\psi}_j\right]$?

- No, because HC standard errors break down (are inconsistent) when # of regressors grow in proportion to sample size.
- Same problem for bootstrap (Bickel and Freedman, 1983)
- To handle high dimensionality: swap usual het-consistent estimators $\hat{\mathbb{V}}_{HC}\left[\hat{\psi}_j\right]$ for het-unbiased estimators $\hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]$. Noise averages out across estimates.

Bias correction: homoscedastic case

Andrews et al (2008): bias correct assuming $\mathbb{V}[\varepsilon] = I\sigma^2$

$$\mathbb{V}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\sigma^2$$

where $\tilde{F}$ is residualized version of F (against D and X).

- Estimate $\mathbb{V}\left[\hat{\psi}\right]$ using DoF adjusted regression MSE

$$\hat{\sigma}^2 = \frac{SSR}{n - \dim(Z)}$$

- But homoscedasticity is a strong assumption
  - Can't be correct if outcome is bounded
  - And in the case of log wages there is ample evidence that error variance differs by gender / experience (e.g., Lemieux, 2006)

Bias correction: heteroscedasticity

Index each person-year observation by $\ell = \ell(i, t)$

- Suppose errors $\{\varepsilon_\ell\}$ are mutually independent
- But potentially heteroscedastic with variances $\sigma_\ell^2 = \mathbb{V}[\varepsilon_\ell]$

Yields familiar "sandwich" variance expression (White, 1980)

$$\mathbb{V}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\Omega\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

where $\Omega = diag\left(\sigma_1^2, ..., \sigma_n^2\right)$.

Estimation challenge: How to get the error variances $\left\{\sigma_\ell^2\right\}_{\ell=1}^n$?

Kline, Saggio, and Sølvsten (2020)

Write AKM as high-dimensional regression:

$$Y_\ell = Z_\ell'\beta + \varepsilon_\ell, \quad \text{for } \ell = 1, ..., n.$$

- Let $\hat{\beta}_{-\ell}$ denote the OLS estimator of β obtained after leaving out obs ℓ. (Requires leave-out connectedness)
- "Cross-fit" estimator of σ_ℓ^2 is *unbiased*:

$$\hat{\sigma}_\ell^2 \quad = \quad Y_\ell \quad \underbrace{\left(Y_\ell - Z_\ell'\hat{\beta}_{-\ell}\right)}_{\text{leave-out prediction error}}$$

Cross-fitting

"Cross-fit" estimator of σ_ℓ^2 is *unbiased*:

$$\begin{aligned}\hat{\sigma}_\ell^2 &= Y_\ell \underbrace{\left(Y_\ell - Z_\ell' \hat{\beta}_{-\ell}\right)}_{\text{leave-out prediction error}} \\ &= \left(\varepsilon_\ell + Z_\ell' \beta\right)\left(\varepsilon_\ell + Z_\ell'\left(\beta - \hat{\beta}_{-\ell}\right)\right)\end{aligned}$$

Intuition: leave-out breaks corr between $\hat{\beta}$ and ε_ℓ

$$\begin{aligned}\mathbb{E}\left[\varepsilon_\ell\left(\beta - \hat{\beta}_{-\ell}\right)\right] &= \mathbb{E}\left[\varepsilon_\ell \left(\sum_{l \neq \ell} Z_l Z_l'\right)^{-1} \sum_{l \neq \ell} Z_l \varepsilon_l\right] \\ &= \left(\sum_{l \neq \ell} Z_l Z_l'\right)^{-1} \sum_{l \neq \ell} Z_l \underbrace{\mathbb{E}\left[\varepsilon_\ell \varepsilon_l\right]}_{=0}\end{aligned}$$

Bias correction

Proxy Ω with $\hat{\Omega} = diag\left\{\hat{\sigma}_\ell^2\right\}_{\ell=1}^n$ to get unbiased variance estimates

$$\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

Bias corrected estimator of θ_ψ is:

$$\hat{\theta}_{\psi,HU} = \underbrace{\hat{\theta}_\psi}_{\text{plugin}} - \underbrace{\sum_{j=1}^{J} s_j \hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]}_{\text{average squared stderr}} + \underbrace{\hat{\mathbb{V}}_{HU}\left[\hat{\bar{\psi}}\right]}_{\text{stderr of mean}}$$

Generalization

What about other variances and covariances?

- KSS consider more general (co-)variance components

$$\theta = \beta' A \beta$$

where A is user specified matrix.

- General bias correction formula:

$$\hat{\theta}_{HU} = \hat{\theta} - \sum_{\ell=1}^{n} B_{\ell\ell} \hat{\sigma}_{\ell}^{2}$$

where $B_{\ell\ell} = Z_{\ell}' \left(\sum_{l=1}^{n} Z_l Z_l'\right)^{-1} A \left(\sum_{l=1}^{n} Z_l Z_l'\right)^{-1} Z_{\ell}$ gives influence of ε_{ℓ}^{2} on $\hat{\theta}$. Mathematical intuition:

$$\hat{\theta} = \theta + \sum_{\ell=1}^{n} B_{\ell\ell} \varepsilon_{\ell}^{2} + o_p(1).$$

Computation

A useful trick:

$$\begin{aligned}\hat{\sigma}_\ell^2 &= Y_\ell \left(Y_\ell - Z_\ell' \hat{\beta}_{-\ell} \right) \\ &= Y_\ell \frac{\left(Y_\ell - Z_\ell' \hat{\beta} \right)}{1 - P_{\ell\ell}}\end{aligned}$$

where $\{P_{\ell\ell}\}$ are the diagonal elements of $P = Z\left(Z'Z\right)^{-1} Z'$.

- Note: only need to compute $\hat{\beta}$ once!
- In large problems can stochastically approximate $\{B_{\ell\ell}, P_{\ell\ell}\}$ (CHK size application in $<$1hr)
- Code / executables available at GitHub repository

Application to Italian data

Administrative records from Italian province of Veneto

Compare plug-in (AKM), homoscedasticity-only (HO) estimator of Andrews (2008), and KSS

Base sample: two wage observations per worker

- With a single wage change per worker we can ignore serial correlation / clustering when computing firm effect variances
- Allows us to focus on importance of heteroscedasticity, but throws away some of the data
- Analyzing 6 year panel via leave-worker-out yields similar results

Split by age: older workers move less $\Rightarrow$ more bias

Bias correction to variance of firm effs

Homoscedastic correction about half way between naive plug-in and KSS

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Large bias in correlation coefficient

Flips sign in age-specific samples!

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Small decrease in total explanatory power of model

Note: HO estimate is familiar “adjusted” R^2, which seems to exhibit negligible bias.

VARIANCE DECOMPOSITION[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug in (PI)	0.0358	0.0368	0.0415
Homoscedasticity Only (HO)	0.0295	0.0270	0.0350
Leave Out (KSS)	0.0240	0.0218	0.0204
Variance of Person Effects			
Plug in (PI)	0.1321	0.0843	0.2180
Homoscedasticity Only (HO)	0.1173	0.0647	0.2046
Leave Out (KSS)	0.1119	0.0596	0.1910
Covariance of Firm, Person Effects			
Plug in (PI)	0.0039	−0.0058	−0.0032
Homoscedasticity Only (HO)	0.0097	0.0030	0.0040
Leave Out (KSS)	0.0147	0.0075	0.0171
Correlation of Firm, Person Effects			
Plug in (PI)	0.0565	−0.1040	−0.0334
Homoscedasticity Only (HO)	0.1649	0.0726	0.0475
Leave Out (KSS)	0.2830	0.2092	0.2744
Coefficient of Determination (R^2)			
Plug in (PI)	0.9546	0.9183	0.9774
Homoscedasticity Only (HO)	0.9029	0.8184	0.9524
Leave Out (KSS)	0.8976	0.8091	0.9489

Estimates from 6 year panel nearly identical after accounting for serial correlation

TABLE A.I

VARIANCE OF FIRM EFFECTS UNDER DIFFERENT LEAVE-OUT STRATEGIES[a]

	Pooled	Younger Workers	Older Workers
Variance of Firm Effects			
Plug-in	0.0304	0.0303	0.0376
Leave Person-Year Out	0.0296	0.0302	0.0314
Leave Match Out	0.0243	0.0221	0.0265
Leave Worker Out	0.0241	0.0227	0.0270

Leaving match out yields same answer as leaving whole worker out
$\Rightarrow$ sufficient to "cluster" std err estimates $\hat{\mathbb{V}}_{HU}\left[\hat{\psi}_j\right]$ by match

Projecting fixed effects onto observables

- Common to project fixed effect estimates $\hat{\psi}$ onto covariates
- Problem: $\hat{\psi}$ are correlated with one another
- Dependence hinges on design because

$$\hat{\psi} = \psi + \underbrace{\left(\tilde{F}'\tilde{F}\right)^{-1}\tilde{F}'\varepsilon}_{\text{correlated noise}}$$

- Solution: use HU variance estimator

$$\begin{aligned}\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] &= \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1} \\ &= \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n}\tilde{f}_\ell\tilde{f}_\ell'\hat{\sigma}_\ell^2\right)\left(\tilde{F}'\tilde{F}\right)^{-1}\end{aligned}$$

Connection to HC2

HC2 estimator (Mackinnon and White, 1985) is:

$$\hat{\mathbb{V}}_{HC2}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n} \tilde{f}_\ell \tilde{f}_\ell' \frac{\left(Y_\ell - Z_\ell'\hat{\beta}\right)^2}{1 - P_{\ell\ell}}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

- HC2 is unbiased under *homo*-scedasticity but otherwise inconsistent when $dim\left(\tilde{F}\right) \propto n$.

HU estimator is:

$$\hat{\mathbb{V}}_{HU}\left[\hat{\psi}\right] = \left(\tilde{F}'\tilde{F}\right)^{-1}\left(\sum_{\ell=1}^{n} \tilde{f}_\ell \tilde{f}_\ell' \frac{Y_\ell\left(Y_\ell - Z_\ell'\hat{\beta}\right)}{1 - P_{\ell\ell}}\right)\left(\tilde{F}'\tilde{F}\right)^{-1}$$

- Unbiased under arbitrary heteroscedasticity.

Standard errors on projection

Projection of ψ onto W is linear combination:

$$\left(W'W\right)^{-1} W'\psi = v'\psi$$

- Estimator of variance of projection coefficients is

$$\hat{\mathbb{V}}_{HU}\left[v'\hat{\psi}\right] = v'\left(\tilde{F}'\tilde{F}\right)^{-1}\left(\tilde{F}'\hat{\Omega}\tilde{F}\right)\left(\tilde{F}'\tilde{F}\right)^{-1} v$$

- Suppose v is $J \times 1$ (i.e., single projection coefficient of interest)
- Provided v' doesn't place "too much" weight on any particular coefficient KSS show that:

$$\frac{v'\left(\hat{\psi}-\psi\right)}{\sqrt{\hat{\mathbb{V}}_{HU}\left[v'\hat{\psi}\right]}} \rightarrow N\left(0,1\right)$$

lincom_KSS: high-dim version of Stata "lincom" command

Naive “robust” std err order of magnitude too small!

PROJECTING FIRM EFFECTS ONTO COVARIATES[a]

	(1)	(2)
Older Worker	0.0272	−0.0016
	(0.0009)	(0.0024)
	[0.0003]	[0.0001]
Log Firm Size		0.0276
		(0.0007)
		[0.0001]
Older Worker × Log Firm Size		0.0028
		(0.0005)
		[0.0002]
Predicted Gap in Firm Effects (Older vs. Younger Workers)	0.0272	0.0054
	(0.0009)	(0.0019)
	[0.0003]	[0.0008]
Number of Observations	1,319,972	1,319,972

[a]This table reports the coefficients from projections of firm effects onto worker and firm characteristics in the pooled leave-one-out sample. A constant is included in each model. Standard errors based on equation (7) reported in parentheses. Naive Eicker–White (HC1) standard errors shown in square brackets. “Predicted Gap in Firm Effects” reports the predicted difference in firm effects between older and younger workers according to either Column (1) or Column (2) evaluated at the median firm size of 12 workers.

Naive std error on old dummy off by a factor of 24 in Col 2! Leave out std error reveals that older workers no more likely to work at high paying firms after adjusting for firm size.

Testing high dimensional hypotheses about fixed effects

Do the firm effects for younger workers equal those faced by older workers?

$$H_0 : \psi_j^O = \psi_j^Y \quad \text{for } j = 1, ..., J$$

- $J = 8,578 \Rightarrow$ cannot rely on standard $\chi^2\,(8578)$ approximation to F-test
- Bootstrap also fails

KSS: test by estimating the variance component

$$\theta_{H_0} = \frac{1}{8578} \left(\psi^O - \psi^Y\right)' \left(\tilde{F}'\tilde{F}\right) \left(\psi^O - \psi^Y\right)$$

Intuition:

- If H_0 is true, we must have $\theta_{H_0} = 0$
- $\tilde{F}'\tilde{F}$ gives optimal (i.e. inverse variance) weighting of differences $\hat{\psi}^O - \hat{\psi}^Y$ under homoscedasticity

Testing high dimensional hypotheses about fixed effects

Do the firm effects for younger workers equal those faced by older workers?

$$H_0 : \psi_j^O = \psi_j^Y \quad \text{for } j = 1, ..., J$$

KSS: test by estimating the variance component

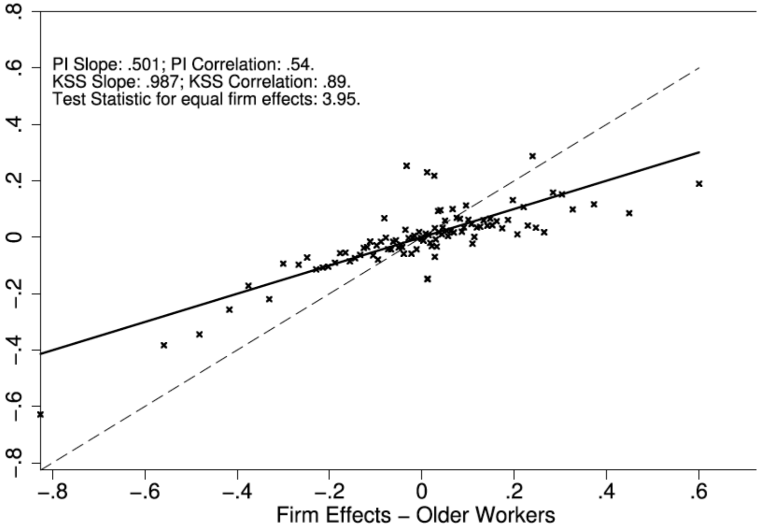
$$\theta_{H_0} = \frac{1}{8578} \left(\psi^O - \psi^Y\right)' \left(\tilde{F}'\tilde{F}\right) \left(\psi^O - \psi^Y\right)$$

Under H_0: $\hat{\theta}_{H_0}$converges to $\mathcal{N}\left(0, \mathbb{V}\left[\hat{\theta}_{H_0}\right]\right)$.

- Estimation of $\mathbb{V}\left[\hat{\theta}_{H_0}\right]$ explained in paper.
- Test statistic is simple t-stat $\hat{\theta}_{H_0} / \sqrt{\hat{\mathbb{V}}_{HU}\left[\hat{\theta}_{H_0}\right]}$

Firm effects highly correlated across age groups

But can decisively reject that they are exactly the same



Gerard, Lagos, Severnini, and Card (2018)

Do different racial groups share equally in firm effects?

Fit AKM model to Brazilian data 2002-2014

- Bias correct via KSS
- Apply high dimensional Oaxaca decomp ala CCK (2016)
- Usual sorting component additionally decomposed based upon regional racial / education shares

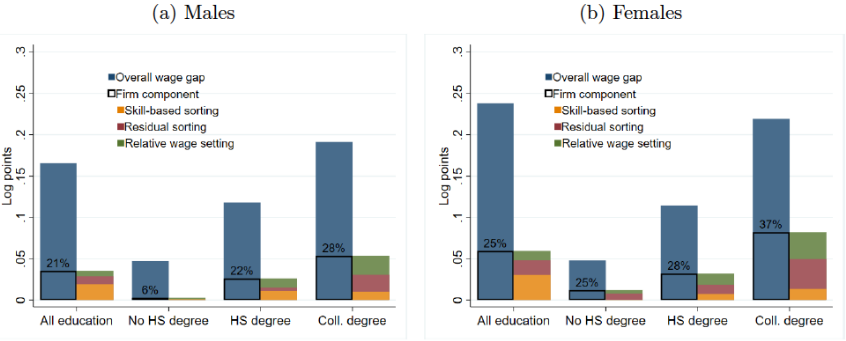
Estab effs ∼14-16% of variance for each group

	White male (1)	Non-white male (2)	White female (3)	Non-white female (4)
Largest connected set				
Standard deviation of log wages	0.670	0.582	0.685	0.554
Mean log wages	1.939	1.768	1.782	1.557
A. AKM decomposition				
Std. dev. of person effects (across person-yr obs.)	0.484	0.415	0.527	0.437
Std. dev. of estab. effects (across person-yr obs.)	0.304	0.279	0.304	0.266
Std. dev. of covariates (across person-yr obs.)	0.175	0.181	0.181	0.185
Correlation of person/estab. effects	0.275	0.167	0.264	0.102
Adjusted R-squared of model	0.901	0.876	0.918	0.897
Percentage of variance of log wages due to:				
person effect	52.1%	50.9%	59.1%	62.1%
establishment effect	20.6%	23.1%	19.7%	23.1%
covariance of person and estab. effects	18.0%	11.4%	18.0%	7.7%
estab. effects+covariance person and estab. effects	38.6%	34.5%	37.7%	30.9%
Number of establishments	1,284,740	717,098	1,162,373	508,088
Number of movers	4,052,299	1,771,840	2,845,495	930,306
Number of person-year observations	39,661,514	16,605,082	27,814,349	8,900,093
Leave-one-out connected set				
Standard deviation of log wages	0.670	0.576	0.706	0.569
Mean log wages	1.961	1.775	1.847	1.582
B. AKM decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.287	0.253	0.293	0.236
Correlation of person/estab. effects	0.354	0.261	0.375	0.260
Percentage of variance of log wages due to:				
establishment effect	18.3%	19.3%	17.2%	17.2%
covariance of person and estab. effects	21.3%	15.7%	23.0%	16.2%
estab. effects+covariance person and estab. effects	39.6%	35.0%	40.2%	33.4%
C. KSS decomposition				
Std. dev. of estab. effects (across person-yr obs.)	0.271	0.233	0.273	0.213
Correlation of person/estab. effects	0.468	0.397	0.480	0.369
Percentage of variance of log wages due to:				
establishment effect	16.3%	16.4%	15.0%	14.0%
covariance of person and estab. effects	22.8%	17.6%	24.6%	18.4%
estab. effects+covariance person and estab. effects	39.1%	34.0%	39.6%	32.5%
Number of establishments	749,877	325,034	600,499	179,697
Number of movers	3,551,977	1,423,252	2,328,539	645,146
Number of person-year observations	22,305,141	8,761,529	13,972,235	3,708,699

Bias corrections to variance shares small w/ 12 years of data

But correction to worker-estab correlation is substantial..

Race gap in estab effs most important for coll educated



Bars give gap between whites and non-whites. Percentages are portion of overall gap attributable to firm components.

Lachowska, Mas, Saggio, Woodbury (2020)

How stable are firm effects?

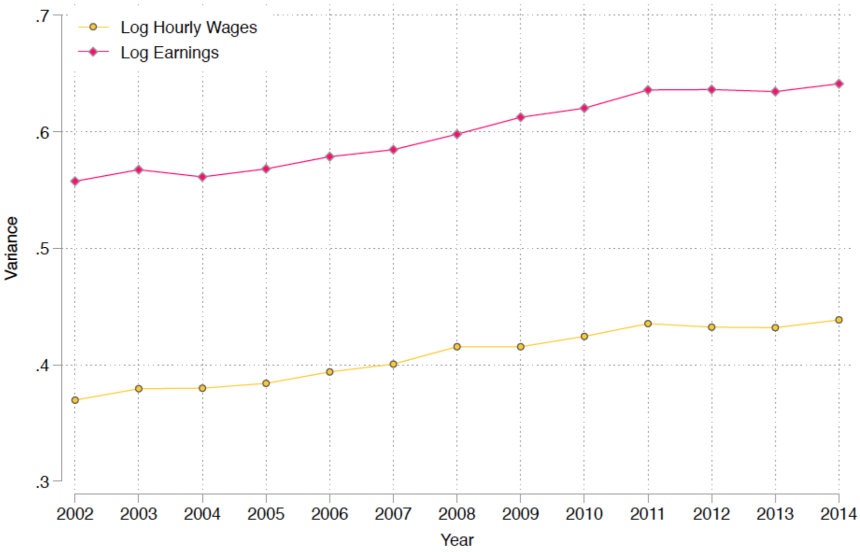
Answer using admin data from Washington state (2002-2014)

- Administrative hours records allow computation of hourly wage
- Secular increase in inequality + Great Recession make for an interesting test environment

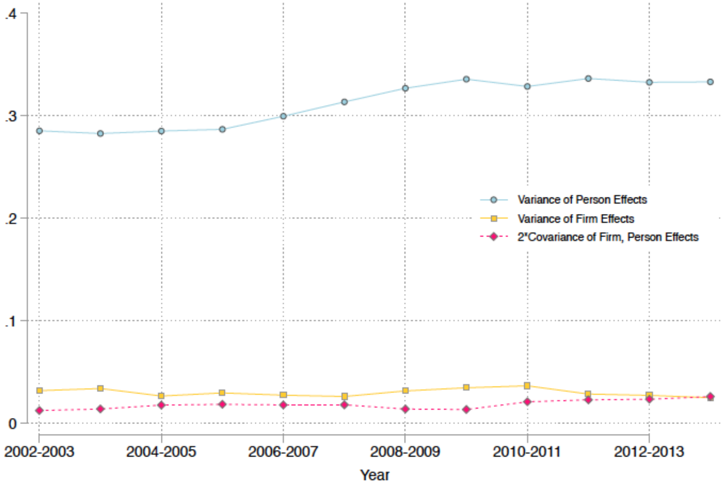
Fit AKM to rolling two year windows of administrative ("TV-AKM")

- Bias correct variance components ala KSS
- Compute autocorrelation of firm effects across windows

Secular increase in log wage variance

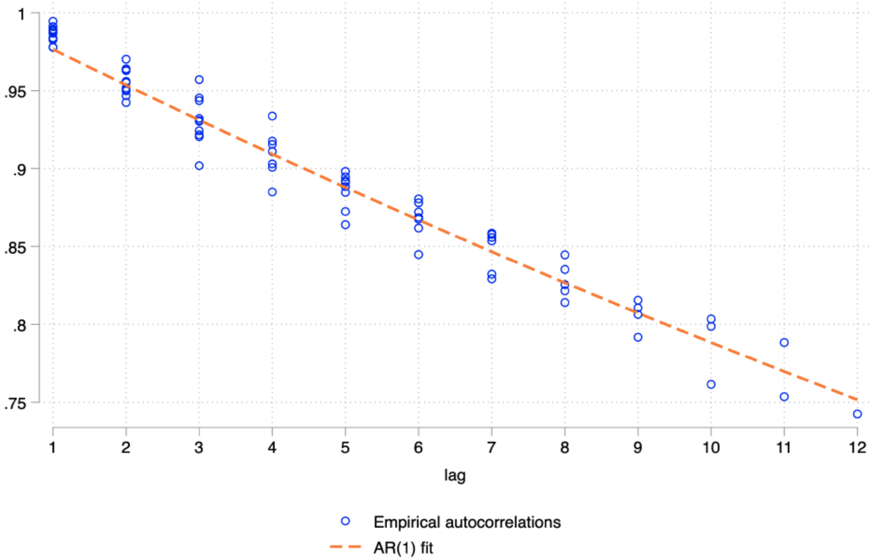


Mostly explained by increase in variance of person effects



Firm effects highly persistent ($\rho \approx .98$)

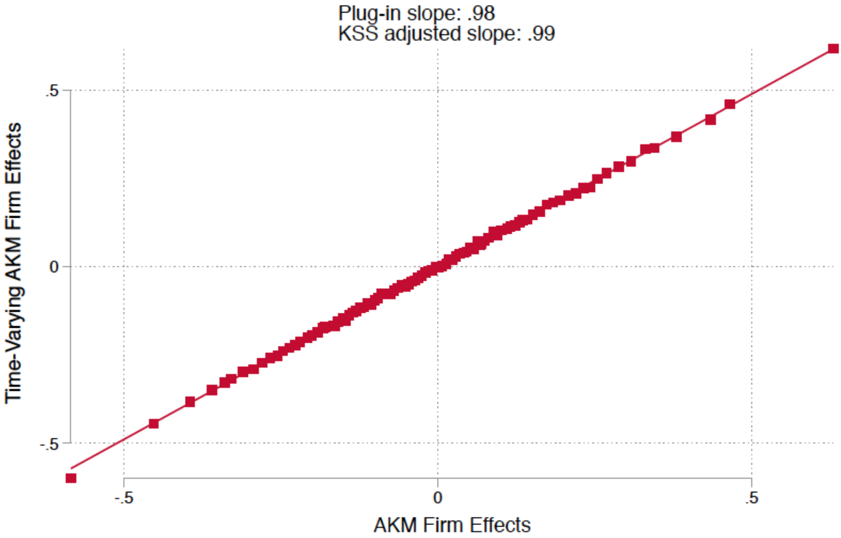
Figure 7: Autocorrelation of Firm Effects for Wages



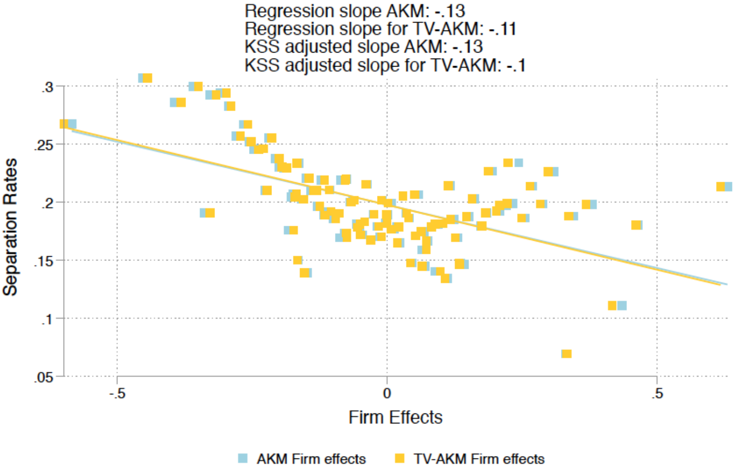
Little bias from imposing constant effects

Figure 9: Does Allowing for Time-Varying Firm Heterogeneity Actually Matter?

(a) TV-AKM vs. AKM



Constant effects also predict separations equally well



Summary

- Statistical firm wage effects temporally stable and correlate strongly with worker retention and productivity
- But not all workers share equally in firm effects
- And "fissuring" the firm via outsourcing leads to wage losses largely explained by firm effects
- Next lecture: What do firm effects tell us about how labor markets actually function?

References

1. Abowd, J. M., Creecy, R. H., & Kramarz, F. (2002). Computing person and firm effects using linked longitudinal employer-employee data (No. 2002-06). Center for Economic Studies, US Census Bureau.
2. Abowd, J. M., Kramarz, F., & Margolis, D. N. (1999). High wage workers and high wage firms. Econometrica, 67(2), 251-333.
3. Abowd, J. M., Lengermann, P. A., & McKinney, K. L. (2003). The Measurement of Human Capital in the US Economy (No. 2002-09). Center for Economic Studies, US Census Bureau.
4. Andrews, M. J., Gill, L., Schank, T., & Upward, R. (2008). High wage workers and low wage firms: negative assortative matching or limited mobility bias?. Journal of the Royal Statistical Society: Series A (Statistics in Society), 171(3), 673-697.
5. Babcock, L., & Laschever, S. (2009). Women don't ask: Negotiation and the gender divide. Princeton University Press.
6. Bertheau, Acabbi, Barcelo, Gulyas, Lombardi and Saggio. 2021. The Costs of Job Loss across Europe: Evidence and Explanations
7. Bickel, P. J., & Freedman, D. A. (1983). Bootstrapping regression models with many parameters. Festschrift for Erich L. Lehmann, 28-48.

8. Card, D., Heining, J., & Kline, P. (2013). Workplace heterogeneity and the rise of West German wage inequality. The Quarterly journal of economics, 128(3), 967-1015.
9. Card, D., Cardoso, A. R., & Kline, P. (2016). Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women. The Quarterly Journal of Economics, 131(2), 633-686.
10. Dustmann, C., Ludsteck, J., & Schönberg, U. (2009). Revisiting the German Wage Structure. The Quarterly Journal of Economics, 124(2), 843-881.
11. Fredriksson, P., Hensvik, L., & Nordstrom Skans, O. (2015). Mismatch of talent: evidence on match quality, entry wages, and job mobility. IZA Working Paper.
12. Freedman, D. A. (1991). Statistical models and shoe leather. Sociological methodology, 291-313.
13. Gerard, F., Lagos, L., Severnini, E., & Card, D. (2018). Assortative matching or exclusionary hiring? the impact of firm policies on racial wage differences in brazil (No. w25176). National Bureau of Economic Research.
14. Goldschmidt, D., & Schmieder, J. F. (2017). The rise of domestic outsourcing and the evolution of the German wage structure. The Quarterly Journal of Economics, 132(3), 1165-1217.
15. Gibbons, R., & Katz, L. (1992). Does unmeasured ability explain inter-industry wage differentials?. The Review of Economic Studies, 59(3), 515-535.

16. Huneeus, C., Neilson, C., Miller, C., & Zimmerman, S. (2015). Firm Matching and the Returns to Elite College Education. Working paper.
17. Kline, P., Saggio, R., & Sølvsten, M. (2020). Leave-out estimation of variance components. Econometrica, 88(5), 1859-1898.
18. Krueger, A. B., & Summers, L. H. (1988). Efficiency wages and the inter-industry wage structure. Econometrica: Journal of the Econometric Society, 259-293.
19. Lachowska, M., Mas, A., Saggio, R. D., & Woodbury, S. A. (2020). Do firm effects drift? Evidence from Washington administrative data (No. w26653). National Bureau of Economic Research.
20. Lachowska, Marta, Alexandre Mas, and Stephen A. Woodbury. 2020. "Sources of Displaced Workers' Long-Term Earnings Losses." American Economic Review, 110 (10): 3231-66.
21. Lemieux, T. (2006). Increasing residual wage inequality: Composition effects, noisy data, or rising demand for skill?. American Economic Review, 96(3), 461-498.
22. MacKinnon, J. G., & White, H. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. Journal of econometrics, 29(3), 305-325.
23. Schmieder, J., von Wachter, T., & Heining, J. (2018). The costs of job displacement over the business cycle and its sources: evidence from Germany. UCLA, Mimeo.

24. Slichter, S. H. (1950). Notes on the Structure of Wages. The Review of Economics and Statistics, 80-91.
25. Weil, D. (2014). The fissured workplace. Harvard University Press.
26. White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. Econometrica: journal of the Econometric Society, 817-838.

AEA Continuing Education 2021: Labor Economics and Applied Econometrics

Lecture #3 - Monopsony, Rent Sharing, and Outside Options

Patrick Kline

UC Berkeley

Monopsony Overview

- Joan Robinson (1933) proposed theory of monopsony in *The Economics of Imperfect Competition*
- Manning (2002) *Monopsony in Motion* argues monopsony more widespread than commonly understood. Recent review in Manning (2020).
- Card, Cardoso, Heining, Kline (2018) contrast perspective of IO literature with that of labor literature:

  Although economists seem to agree that part of the variation in the prices of cars and breakfast cereal is due to factors other than marginal cost, the notion that wages differ substantially among equally skilled workers remains highly controversial.

Card, Cardoso, Heining, Kline (2018)

Borrow from IO literature on differentiated product markets

- Basic idea: firms imperfect substitutes in eyes of workers
- Endows firm with power to set wages

Study conditions under which stable firm effects arise along with their interpretation

Link to older empirical literature on rent sharing

Setup

- Two observable worker types $S \in \{L, H\}$ w/ corresponding market supplies $(\mathcal{L}, \mathcal{H})$
- J firms, differentiated vertically by amenities a_{Sj}
- Worker-firm pairings yield match effects $\{\epsilon_{iSj}\}$ that are *private information* to workers

Indirect utility of working at firm j for skill type S:

$$u_{iSj} = \underbrace{\beta_S \ln(w_{Sj} - b_S)}_{\text{wages}} + \underbrace{a_{Sj}}_{\text{amenities}} + \underbrace{\epsilon_{iSj}}_{\text{match}}$$

Here b_S is a type-specific reservation wage / outside option

- Analogous to Stone-Geary min consumption level
- Will not work for less, no matter amenity level.

Labor supply to firm

Assuming $\epsilon_{iSj} \sim EVI$ we have LS curves

$$\begin{aligned} \ln L_j\left(w_{Lj}\right) &= \ln\left(\mathcal{L}\lambda_L\right) + \beta_L \ln\left(w_{Lj} - b_L\right) + a_{Lj} \\ \ln H_j\left(w_{Hj}\right) &= \ln\left(\mathcal{H}\lambda_H\right) + \beta_H \ln\left(w_{Hj} - b_H\right) + a_{Hj} \end{aligned}$$

where $\lambda_S \equiv \sum_{k=1}^{J} \exp\left(\beta_S \ln\left(w_{Sk} - b_S\right) + a_{Sk}\right)$

Labor supply to firm

Assuming $\epsilon_{iSj} \sim EVI$ we have LS curves

$$
\begin{aligned}
\ln L_j\left(w_{Lj}\right) &= \ln\left(\mathcal{L}\lambda_L\right) + \beta_L \ln\left(w_{Lj} - b_L\right) + a_{Lj} \\
\ln H_j\left(w_{Hj}\right) &= \ln\left(\mathcal{H}\lambda_H\right) + \beta_H \ln\left(w_{Hj} - b_H\right) + a_{Hj}
\end{aligned}
$$

where $\lambda_S \equiv \sum_{k=1}^{J} \exp\left(\beta_S \ln\left(w_{Sk} - b_S\right) + a_{Sk}\right)$

Supposing J is large can approximate λ_S as constant

Labor supply to firm

Assuming $\epsilon_{iSj} \sim EVI$ we have LS curves

$$\begin{aligned} \ln L_j\left(w_{Lj}\right) &= \ln\left(\mathcal{L}\lambda_L\right) + \beta_L \ln\left(w_{Lj} - b_L\right) + a_{Lj} \\ \ln H_j\left(w_{Hj}\right) &= \ln\left(\mathcal{H}\lambda_H\right) + \beta_H \ln\left(w_{Hj} - b_H\right) + a_{Hj} \end{aligned}$$

where $\lambda_S \equiv \sum_{k=1}^{J} \exp\left(\beta_S \ln\left(w_{Sk} - b_S\right) + a_{Sk}\right)$

Supposing J is large can approximate λ_S as constant

Approximation yields variable LS elasticity:

$$e_{Sj} = \beta_S \cdot \frac{w_{Sj}}{w_{Sj} - b_S}$$

- Decreasing in w_{Sj} (infinite at $w_{Sj} = b_S$)
- Approach competitive model as $\beta_S \to \infty$

Firm's problem

Firm j's output given by

$$Y_j = T_j f\left(L_j, H_j\right)$$

where T_j is TFPQ and $f\left(.,.\right)$ is CRTS fn

Firm's problem

Firm j's output given by

$$Y_j = T_j f(L_j, H_j)$$

where T_j is TFPQ and $f(.,.)$ is CRTS fn

Choose wages to minimize costs subject to output $\geq Y$

$$\min_{w_{Lj}, w_{Hj}} w_{Lj} L_j(w_{Lj}) + w_{Hj} H_j(w_{Hj})$$

$$\text{s.t.}\, T_j f(L_j(w_{Lj}), H_j(w_{Hj})) \geq Y$$

- Firm wage discriminates on S but not ϵ_{iSj} (2nd degree)
- Large market approximation: ignores effect of wage choice on behavior of other firms

Wage rule

FOC yields monopsony "markdown" rule (Robinson, 1933):

$$w_{Sj} = \underbrace{\frac{e_{Sj}}{1 + e_{Sj}}}_{\text{exploitation}} \underbrace{T_j f_S \mu_j}_{\text{MRPL}}$$

- μ_j is Lagrange multiplier on output constraint
- Firm will choose output to equate MC (μ_j) with MR
- Wage sets MFC $\left(\frac{1+e_{Sj}}{e_{Sj}}\right) w_{Sj}$ equal to MRPL

Wage rule

FOC yields monopsony "markdown" rule (Robinson, 1933):

$$w_{Sj} = \underbrace{\frac{e_{Sj}}{1+e_{Sj}}}_{\text{exploitation}} \underbrace{T_j f_S \mu_j}_{\text{MRPL}}$$

- μ_j is Lagrange multiplier on output constraint
- Firm will choose output to equate MC (μ_j) with MR
- Wage sets MFC $\left(\frac{1+e_{Sj}}{e_{Sj}}\right) w_{Sj}$ equal to MRPL

Using $e_{Sj} = \frac{\beta_S w_{Sj}}{w_{Sj} - b_S}$ we get

$$w_{Sj} = \frac{1}{1+\beta_S} b_S + \frac{\beta_S}{1+\beta_S} MRPL_j$$

- Wage is a weighted average of outside option and MRPL
- $\frac{\beta_S}{1+\beta_S}$ analogous to Nash bargaining weight

A Baseline Case

Linear production (efficiency units)

$$Y_j = T_j \left[(1-\theta) L_j + \theta H_j \right] \equiv T_j N_j$$

With fixed product price P_j^0, value added per eff unit of labor is:

$$v_j \equiv P_j^0 Y_j / N_j = P_j^0 T_j$$

A Baseline Case

Linear production (efficiency units)

$$Y_j = T_j \left[(1-\theta) L_j + \theta H_j \right] \equiv T_j N_j$$

With fixed product price P_j^0, value added per eff unit of labor is:

$$v_j \equiv P_j^0 Y_j / N_j = P_j^0 T_j$$

Wages are linear in v_j:

$$\begin{aligned} w_{Lj} &= \frac{1}{1+\beta_L} b_L + \frac{\beta_L}{1+\beta_L} (1-\theta) v_j \\ w_{Hj} &= \frac{1}{1+\beta_H} b_H + \frac{\beta_H}{1+\beta_H} \theta v_j \end{aligned}$$

A Baseline Case

Linear production (efficiency units)

$$Y_j = T_j \left[(1-\theta) L_j + \theta H_j \right] \equiv T_j N_j$$

With fixed product price P_j^0, value added per eff unit of labor is:

$$v_j \equiv P_j^0 Y_j / N_j = P_j^0 T_j$$

Wages are linear in v_j:

$$\begin{aligned} w_{Lj} &= \frac{1}{1+\beta_L} b_L + \frac{\beta_L}{1+\beta_L} (1-\theta) v_j \\ w_{Hj} &= \frac{1}{1+\beta_H} b_H + \frac{\beta_H}{1+\beta_H} \theta v_j \end{aligned}$$

Letting $s_H = \frac{H_j}{L_J + H_j}$, empirical studies typically use

$$\tilde{v}_j = P_j^0 Y_j / (L_j + H_j) = v_j \left[(1-\theta) + 2\theta s_H \right]$$

Elasticities

Suppose reservation wages determined by pay in "competitive fringe" sector that pays b per eff unit, so that

$$b_L = (1-\theta)\, b, \quad b_H = \theta b$$

Log wages become

$$\begin{aligned} \ln w_{Lj} &= \ln \frac{(1-\theta)\, b}{1+\beta_L} + \ln(1+\beta_L R_j) \\ \ln w_{Hj} &= \ln \frac{\theta b}{1+\beta_H} + \ln(1+\beta_H R_j) \end{aligned}$$

where $R_j = v_j / b$ gives ratio of j's labor prod relative to competitive fringe

Elasticities

Suppose reservation wages determined by pay in "competitive fringe" sector that pays b per eff unit, so that

$$b_L = (1-\theta)\, b, \quad b_H = \theta b$$

Log wages become

$$\begin{aligned} \ln w_{Lj} &= \ln \frac{(1-\theta)\, b}{1+\beta_L} + \ln (1 + \beta_L R_j) \\ \ln w_{Hj} &= \ln \frac{\theta b}{1+\beta_H} + \ln (1 + \beta_H R_j) \end{aligned}$$

where $R_j = v_j / b$ gives ratio of j's labor prod relative to competitive fringe

Potentially type-specific "rent sharing" elasticity

$$\xi_j \equiv \frac{d \ln w_{Sj}}{d \ln v_j} = \frac{\beta_S R_j}{1 + \beta_S R_j}$$

Three generations of rent-sharing elasticities

Group 1: Industry-level profit measure	
Christofides-Oswald (QJE 1992), Canadian manufacturing	0.140 (0.035)
Blanchflower-Oswald-Sanfey (QJE 1996), US manufacturing	0.060 (0.024)
Group 2: Firm-level profit measure, mean firm wage	
Abowd-Lemieux (QJE 1993), Canadian manufacturing	0.220 (0.081)
Van Reenen (QJE 1996), UK manufacturing	0.290 (0.089)
Barth-Bryson-Davis-Freeman (JOLE 2016), US	0.160 (0.002)
Group 3: Firm-level profit measure, individual-specific wage	
Guiso-Pistaferri-Schivardi (JPE 2005), Italy	0.069 (0.025)
Card-Devicienti-Maida (ReStud 2014), Italy	0.073 (0.031)
Card-Cardoso-Kline (QJE 2014), Portugal, between firm	0.156 (0.006)
Card-Cardoso-Kline (QJE 2014), Portugal, stayers	0.049 (0.007)
Bagger-Christensen-Mortensen (mimeo), Danish manufacturing	0.090 (0.020)

A calibration

$$\xi_j \equiv \frac{d \ln w_{Sj}}{d \ln v_j} = \frac{\beta_S R_j}{1 + \beta_S R_j}$$

Modern estimates give $\xi_j \approx 0.1 \Rightarrow \beta_S R_j \approx 0.1$

Suppose $e_S \approx 4$ (20% markdown), then

- $R_j \approx 1.3$ (30% more productive than competitive fringe)
- $\beta_S \approx 0.08$ (workers get 8 cents of every dollar of MRP)

A link to AKM

When $\beta_L = \beta_H$, we have the AKM representation

$$\ln w_{Lj} = \underbrace{\ln \frac{(1-\theta)\, b}{1+\beta}}_{\alpha_L} + \underbrace{\ln\left(1+\beta R_j\right)}_{\psi_j}$$

$$\ln w_{Hj} = \underbrace{\ln \frac{\theta b}{1+\beta}}_{\alpha_H} + \underbrace{\ln\left(1+\beta R_j\right)}_{\psi_j}$$

A link to AKM

When $\beta_L = \beta_H$, we have the AKM representation

$$\begin{aligned}
\ln w_{Lj} &= \underbrace{\ln \frac{(1-\theta)\, b}{1+\beta}}_{\alpha_L} + \underbrace{\ln (1 + \beta R_j)}_{\psi_j} \\
\ln w_{Hj} &= \underbrace{\ln \frac{\theta b}{1+\beta}}_{\alpha_H} + \underbrace{\ln (1 + \beta R_j)}_{\psi_j}
\end{aligned}$$

For small βR_j, firm effects nearly linear in productivity

$$\psi_j \approx \beta R_j$$

A link to AKM

When $\beta_L = \beta_H$, we have the AKM representation

$$\begin{aligned}
\ln w_{Lj} &= \underbrace{\ln \frac{(1-\theta)\, b}{1+\beta}}_{\alpha_L} + \underbrace{\ln\left(1+\beta R_j\right)}_{\psi_j} \\
\ln w_{Hj} &= \underbrace{\ln \frac{\theta b}{1+\beta}}_{\alpha_H} + \underbrace{\ln\left(1+\beta R_j\right)}_{\psi_j}
\end{aligned}$$

For small βR_j, firm effects nearly linear in productivity

$$\psi_j \approx \beta R_j$$

Limitations

- Firm profits derived entirely from labor market
- Amenities have no effect on ψ_j

Downward sloping product demand

Suppose $P_j = P_j^0 Y_j^{-1/\varepsilon}$ where $\varepsilon > 1$ gives elasticity of demand

- Now avg labor productivity is decreasing in scale

$$v_j = \frac{P_j Y_j}{N_j} = T_j P_j = T_j P_j^0 Y_j^{-1/\varepsilon}$$

- Monopoly rents: mark P_j up over μ_j by a factor $\frac{\varepsilon}{\varepsilon - 1}$

Downward sloping product demand

Suppose $P_j = P_j^0 Y_j^{-1/\varepsilon}$ where $\varepsilon > 1$ gives elasticity of demand

- Now avg labor productivity is decreasing in scale

$$v_j = \frac{P_j Y_j}{N_j} = T_j P_j = T_j P_j^0 Y_j^{-1/\varepsilon}$$

- Monopoly rents: mark P_j up over μ_j by a factor $\frac{\varepsilon}{\varepsilon-1}$

Setting $\mu_j = \left(1 - \frac{1}{\varepsilon}\right) P_j$ we get

$$\begin{aligned} w_{Lj} &= \frac{b(1-\theta)}{1+\beta_L}\left[1 + \beta_L \left(\frac{\varepsilon - 1}{\varepsilon}\right) v_j / b\right] \\ w_{Hj} &= \frac{b\theta}{1+\beta_H}\left[1 + \beta_H \left(\frac{\varepsilon - 1}{\varepsilon}\right) v_j / b\right] \end{aligned}$$

- $\left(\frac{\varepsilon-1}{\varepsilon}\right)$ converts avg to marginal labor productivity
- Amenities affect wages indirectly *through* v_j

Link to AKM

Suppose $\beta_H = \beta_L = \beta$ and take logs to get

$$\ln w_{Lj} = \underbrace{\ln \frac{b(1-\theta)}{1+\beta}}_{\alpha_L} + \underbrace{\ln\left[1+\beta R_j'\right]}_{\psi_j}$$

$$\ln w_{Hj} = \underbrace{\ln \frac{b\theta}{1+\beta}}_{\alpha_H} + \underbrace{\ln\left[1+\beta R_j'\right]}_{\psi_j}$$

- $R_j' \equiv \left(\frac{\varepsilon-1}{\varepsilon}\right) v_j/b$ is ratio of *marginal* labor productivity to productivity in competitive fringe
- Firm effects explainable by labor productivity b/c amenities only shift intercept (rather than slope) of LS curve

Rent sharing

Wage elasticity wrt value added is:

$$\xi_{Sj} = \frac{d \ln w_{Sj}}{d \ln v_j} = \frac{\beta_S R'_j}{1 + \beta_S R'_j}$$

Rent sharing

Wage elasticity wrt value added is:

$$\xi_{Sj} = \frac{d \ln w_{Sj}}{d \ln v_j} = \frac{\beta_S R'_j}{1 + \beta_S R'_j}$$

Letting $m_j \equiv \frac{d \ln N_j}{d \ln v_j}$, we expect somewhat smaller wage responses to TFPQ shocks than to TFPR

$$\begin{aligned} \frac{d \ln w_{Sj}}{d \ln P^0_j} &= \frac{\varepsilon}{\varepsilon + m_j} \xi_{Sj} \\ \frac{d \ln w_{Sj}}{d \ln T_j} &= \frac{\varepsilon - 1}{\varepsilon + m_j} \xi_{Sj} \end{aligned}$$

A simplified example

- Suppose a single labor type L of measure 1
- Set $b = 0$ so that LS exhibits constant elasticity β

A simplified example

- Suppose a single labor type L of measure 1
- Set $b = 0$ so that LS exhibits constant elasticity β
- Production is $Y_j = T_j L_j = T_j \exp\left(\beta \ln w_j + a_j\right)$

Corresponding wage rule is:

$$w_j \quad = \quad \frac{\beta}{1+\beta} MRP = \frac{\beta}{1+\beta} \frac{\varepsilon - 1}{\varepsilon} T_j P_j^0 Y_j^{-1/\varepsilon}$$

A simplified example

- Suppose a single labor type L of measure 1
- Set $b = 0$ so that LS exhibits constant elasticity β
- Production is $Y_j = T_j L_j = T_j \exp(\beta \ln w_j + a_j)$

Corresponding wage rule is:

$$w_j \quad = \quad \frac{\beta}{1+\beta} MRP = \frac{\beta}{1+\beta} \frac{\varepsilon - 1}{\varepsilon} T_j P_j^0 Y_j^{-1/\varepsilon}$$

Solve out for reduced form

$$\begin{aligned}
\ln w_j &= \ln \left[\frac{\beta}{1+\beta} \frac{\varepsilon - 1}{\varepsilon} \right] + \ln P_j^0 T_j - \frac{1}{\varepsilon} \ln Y_j \\
&= \ln \left[\frac{\beta}{1+\beta} \frac{\varepsilon - 1}{\varepsilon} \right] + \ln P_j^0 T_j - \frac{1}{\varepsilon} \left[\ln T_j + \beta \ln w_j + a_j \right] \\
&= \text{constant} + \underbrace{\frac{\varepsilon}{\varepsilon + \beta} \ln P_j^0}_{\text{prod demand}} + \underbrace{\frac{\varepsilon - 1}{\varepsilon + \beta} \ln T_j}_{\text{TFPQ}} - \underbrace{\frac{1}{\varepsilon + \beta} a_j}_{\text{amenities}}
\end{aligned}$$

Interpretation: supply and demand at the firm level

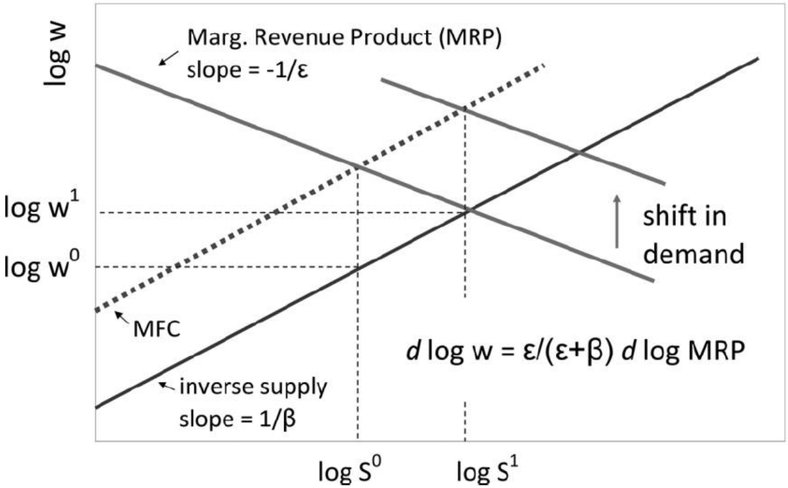


FIG. 9.—Effect of total factor productivity shock (single skill group). MFC = marginal factor cost. A color version of this figure is available online.

Note: S refers to L on previous slide

Summary

Simple "differentiated workplaces" foundation for monopsony easily adapted to many empirical settings

- Forges a link between AKM effects and pass through of productivity shocks to wages
- Microfoundation for firm level supply - demand analysis / study of rents

Summary

Simple “differentiated workplaces” foundation for monopsony easily adapted to many empirical settings

- Forges a link between AKM effects and pass through of productivity shocks to wages
- Microfoundation for firm level supply - demand analysis / study of rents

Extensions

- Imperfect substitution / task assignment at firm level (Haanwinckel, 2018; Lindner et al, 2019)
- Interactions with min wage / other institutions (Haanwinckel, 2018; Berger, Herkenhoff, Mongey, 2019)

Azar, Berry, Marinescu (2019)

Fit differentiated workplace model of LS to online job postings from CareerBuilder.com

- Follow closely standard approaches in empirical IO (e.g., Berry, 1994, BLP, 1996)

Advantages of studying CB

- Posted wages
- Observed application behavior (instead of just realized matches)
- Low search costs on platform

Challenges:

- How to convert application elasticities to LS elasticities?
- Finding exogeneous variation in wages

Nested logit model

Break job vacancies into markets m defined by occupation by geography cells (SOC-6 $\times$ CZ)

Indirect utility of worker i applying to job vacancy $j \in J_{mt}$ in market m in week t is:

$$u_{ijmt} = \delta_j + \gamma_m z_{ijm} + \theta_m \tilde{z}_{im} + \nu_{imt}(\lambda_m) + \lambda_m \epsilon_{ijmt}$$

- δ_j - "mean utility" of job j (treat as fixed effect)
- z_{ijm} - log distance of i to job j
- $\tilde{z}_{im}$ - indicator for i in same CZ as j
- $\nu_{imt}(\lambda_m)$ - market random effect with scale parameter λ_m
- ϵ_{ijmt} - idiosyncratic match $\sim$EV1
- Outside option: don't apply ($j = 0$)

Mean utility

Mean utilities obey:

$$\delta_j = \beta x_j - \alpha \ln w_j + \xi_j$$

- x_j - job characteristics
- w_j - posted wage
- ξ_j - unobserved job "quality"

$Cov\left(\ln w_j, \xi_j\right) > 0 \Rightarrow$ omitted variable bias

Ideal instruments for $\ln w_j$:

- productivity shock
- change in market structure (e.g., merger / outsourcing event)

Nested logit estimation via two-step

Bottom level: choosing jobs within a market

Probability of applying to job j conditional on choosing at least 1 job in market m

$$\begin{aligned} s_{ijmt} &= \frac{\exp\left[\left(\delta_j + \gamma_m z_{ijm}\right)/\lambda_m\right]}{\sum_{k \in J_{mt}} \exp\left[\left(\delta_k + \gamma_m z_{ikm}\right)/\lambda_m\right]} \\ &= \exp\left[\left(\delta_j + \gamma_m z_{ijm}\right)/\lambda_m\right] / \underbrace{I_{imt}}_{\text{inclusive value}} \end{aligned}$$

Can be estimated via conventional alternative specific logit using within market data

Yields scaled mean utilities δ_j/λ_m

Recovering the scale parameter

Top level: which (if any) market to enter

Probability of applying to market m is

$$s_{imt} = \frac{\exp\left(\theta_m \tilde{z}_{im} + \lambda_m I_{imt}\right)}{1 + \exp\left(\theta_m \tilde{z}_{im} + \lambda_m I_{imt}\right)}.$$

Estimate via another logit. Recover scale parameters λ_m. Use to form estimates $\hat{\delta}_j$ of δ_j.

Final step: IV

Explore two sets of instruments for $\ln w_j$ in final equation

$$\hat{\delta}_j = \beta x_j - \alpha \ln w_j + \xi_j + \text{noise}$$

"BLP instruments": # of vacancies in market / size of other firms in market

"Hausman instruments": wages paid by same firm in other markets

- Problem: what if firm wage in other markets reflects unobserved amenities?
- Solution: use predicted wage in other markets (based on CZ-SOC fixed effects + job title fixed effects)
- Intuition: firms that face stiffer competition in other markets it also pays higher wages in this market

Instrumenting flips the sign of wage

But parameter estimates somewhat sensitive to instrument set

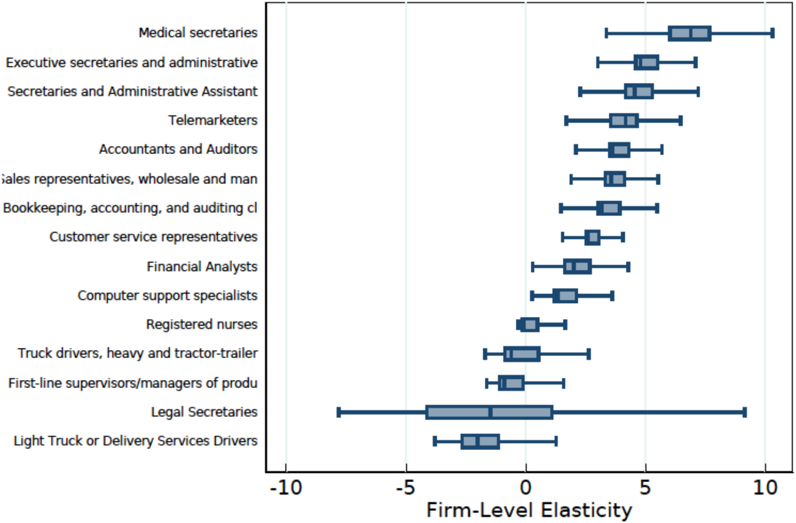
	Dependent variable: δ_j, job-specific utility					
	OLS		IV: Number of Vacancies		IV: BLP Instruments	
	(1)	(2)	(3)	(4)	(5)	(6)
Log Wage	-0.0163***	0.0194***	0.887***	1.316***	0.647***	0.352***
	(0.00387)	(0.00600)	(0.186)	(0.449)	(0.133)	(0.115)
Log Employees	0.00146	0.000451	-0.0734***	-0.0998***	-0.0535***	-0.0253***
	(0.00159)	(0.00198)	(0.0161)	(0.0338)	(0.0115)	(0.00936)
(Log Employees)^2	-0.000710***	-0.000728***	0.00459***	0.00596***	0.00318***	0.000990
	(0.000128)	(0.000162)	(0.00114)	(0.00227)	(0.000823)	(0.000633)
CZ × SOC FE	✓	✓	✓	✓	✓	✓
Job Title FE		✓		✓		✓
Observations	16,481	12,139	16,481	12,139	16,481	12,139
R-squared	0.044	0.052	-4.569	-6.747	-2.439	-0.396
Median Market-Level Elasticity	-0.0135	0.0161	0.734	1.088	0.535	0.291
Median Firm-Level Elasticity	-0.144	0.172	7.839	11.63	5.712	3.112
Median Vacancy-Level Elasticity	-0.147	0.176	8.017	11.89	5.843	3.183
Kleibergen-Paap F-stat			33.69	13.35	17.70	11.92

Note: Vacancy level elasticity > firm level > market level

Hausman instrument somewhat yield lower elasticities

	Dependent variable: δ_j, job-specific utility							
	IV: Average Wage of Same Firm in Other Markets		IV: Average Wage of Same Firm in Other Markets (Excluding Same CZ and Same SOC)		IV: Average Predicted Wage of Same Firm in Other Markets		IV: Average Predicted Wage of Same Firm in Other Markets (Excluding Same CZ and Same SOC)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Wage	0.117***	0.210***	0.148***	0.231***	0.254***	0.528***	0.314***	0.619***
	(0.0178)	(0.0355)	(0.0277)	(0.0540)	(0.0343)	(0.0803)	(0.0470)	(0.122)
Log Employees	-0.0125***	-0.0184***	-0.0205***	-0.0250***	-0.0258***	-0.0466***	-0.0391***	-0.0639***
	(0.00248)	(0.00377)	(0.00373)	(0.00604)	(0.00386)	(0.00751)	(0.00565)	(0.0122)
$(\text{Log Employees})^2$	0.000306	0.000569**	0.000893***	0.00103***	0.00121***	0.00240***	0.00210***	0.00344***
	(0.000187)	(0.000270)	(0.000262)	(0.000398)	(0.000276)	(0.000508)	(0.000386)	(0.000779)
CZ × SOC FE	✓	✓	✓	✓	✓	✓	✓	✓
Job Title FE		✓		✓		✓		✓
Observations	13,865	10,368	11,781	8,851	13,860	10,366	11,777	8,849
R-squared	-0.054	-0.072	-0.108	-0.111	-0.345	-0.823	-0.534	-1.155
Median Market-Level Elasticity	0.0969	0.173	0.122	0.191	0.210	0.437	0.260	0.512
Median Firm-Level Elasticity	1.035	1.852	1.308	2.042	2.241	4.667	2.773	5.471
Median Vacancy-Level Elasticity	1.058	1.895	1.337	2.089	2.292	4.774	2.836	5.596
Kleibergen-Paap F-stat	508.3	146.3	186.7	52.92	229.3	82.36	157.3	43.78

Lots of heterogeneity across occupations



Nurses and truckers are occs that have long been suspected of being monopsonistic (Rose, 1987; Staiger, Spetz, Phibbs, 2010)

Application elasticities to LS elasticities

Definitions:

- $R(w)$ is the flow of new recruits as a function of wage
- $s(w)$ is the separation rate

Steady state: $s(w)\,N(w) = R(w) \Rightarrow N(w) = R(w)/s(w)$

Application elasticities to LS elasticities

Definitions:

- $R(w)$ is the flow of new recruits as a function of wage
- $s(w)$ is the separation rate

Steady state: $s(w)\,N(w) = R(w) \Rightarrow N(w) = R(w)\,/s(w)$

SS elasticity of LS: $\epsilon = \frac{d \ln N}{d \ln w} = \epsilon_R - \epsilon_s$.

Application elasticities to LS elasticities

Definitions:

- $R(w)$ is the flow of new recruits as a function of wage
- $s(w)$ is the separation rate

Steady state: $s(w)\,N(w) = R(w) \Rightarrow N(w) = R(w)/s(w)$

SS elasticity of LS: $\epsilon = \frac{d\ln N}{d\ln w} = \epsilon_R - \epsilon_s$.

Two (strong) assumptions:

1. app elasticity $\approx \epsilon_R$
2. $\epsilon_R \approx \epsilon_s$

(1) + (2) $\Rightarrow \epsilon \approx 2\times$app elasticity

Summary

Standard IO tools can be applied to study labor market competition

- Results somewhat sensitive to instrument set
- Doubling the app elasticity is a crude way to assess the full LS elasticity

Extensions:

- How best to define labor markets? (Manning and Petrongolo 2017; Nimczik, 2017; Caldwell and Daniele, 2018)
- Direct evidence (e.g., mergers / firm entry) on effects of changes in market structure (Arnold, 2019; Manelici and Vasquez, 2019)

Dube, Jacobs, Naidu, Suri (2020)

How much market power do employers have on online labor markets?

Study relationship between reward and availability for Mturk tasks to estimate labor supply curve

- use double machine learning (DML) procedure of Chernozhukov et al (2018) to infer causality
- validate with experiments

Main result: labor supply elasticity to "requester" is very low

A toy model

Requester posts batch of N jobs with private value p that need to be completed in time interval $[0, T]$.

- A fraction λ of users see the request
- Distribution of reservation wages is $F(w)$

Requester chooses a wage to maximize

$$\Pi(w) = \int_0^T e^{-rt} N(w, t)(p - w) F(w) \lambda dt$$

where $N(w, t)$ is the stock of unfilled jobs, which evolves according to

$$\dot{N}(w, t) = -\lambda F(w) N(w, t).$$

Duration elasticity

$$\dot{N}(w,t) = -\lambda F(w) N(w,t).$$

With constant fill rate $\lambda F(w)$, expected duration to fill N jobs (ignoring censoring at T) is

$$\bar{d} = \frac{N}{\lambda F(w)}$$

Imposing $F(w) \propto w^{\eta}$, we have

$$\begin{aligned} \ln \bar{d} &= \ln N - \ln \lambda F(w) \\ &\propto \ln N - \ln \lambda - \eta \ln w \end{aligned}$$

Quasi-static interpretation

For short T, effective LS curve $L(w)$ is

$$L(w) \propto F(w)$$

Elasticity of labor supply equals duration elasticity

$$\frac{d \ln L}{d \ln w} = \frac{d \ln F}{d \ln w} = \frac{d \ln \bar{d}}{d \ln w} = \eta.$$

Econometric framework

MTurk data consist of a series of scraped human input task batches (HITs). Relationship of interest is:

$$\ln\left(duration_h\right) = -\eta \ln\left(reward_h\right) + \nu_h + \epsilon_h$$

- $duration_h$ is the time it took for the HIT to disappear from Mturk
- $reward_h$ is the payment for completing the HIT
- ν_h confounders

η is duration elasticity

- Frictionless competitive model $\eta = \infty$
- Is this a reasonable benchmark?

Panel data estimator

$$\ln\left(\mathit{duration}_h\right) = -\eta \ln\left(\mathit{reward}_h\right) + \nu_h + \epsilon_h$$

Fixed effects for confounders

$$\nu_h = \underbrace{\rho_{r(h)}}_{\text{employer}} + \underbrace{\tau_{t(h)}}_{\text{day}} + \underbrace{\delta_{d(h)}}_{\text{minutes allotted}} + \underbrace{\delta_{N(h)}}_{\text{batch size}}$$

Estimate η by OLS

DML estimator

Partially linear model:

$$\begin{aligned} \ln\left(duration_h\right) &= -\eta \ln\left(reward_h\right) + g_0\left(Z_h\right) + \epsilon_h \\ \ln\left(reward_h\right) &= m_0\left(Z_h\right) + \mu_h \end{aligned}$$

where Z_h is high dimensional vector of HIT features.

1. Estimate first stage function $m_0\left(Z_h\right)$ and reduced form

$$l_0\left(Z_h\right) = \mathbb{E}\left[\ln\left(duration_h\right)|Z_h\right] = g_0\left(Z_h\right) + m_0\left(Z_h\right)$$

   via random forest procedure utilizing classification trees (Breiman, 2001).
2. Form resids:

$$\hat{\xi}_h = \ln\left(duration_h\right) - \hat{l}_0\left(Z_h\right)$$

$$\hat{\mu}_h = \ln\left(reward_h\right) - \hat{m}_0\left(Z_h\right)$$

DML estimator

Frisch-Waugh style estimator of η based on residuals:

$$\hat{\eta} = \left(\sum_h \hat{\mu}_h^2 \right)^{-1} \sum_h \hat{\xi}_h \hat{\mu}_h$$

- Problem: model selection errors in $\hat{\xi}_h$ and $\hat{\mu}_h$ could be correlated, amplifying regularization bias
- Solution: split sample to obtain independent $\hat{\xi}_h^{(1)}$ and $\hat{\mu}_h^{(2)}$

Chernozhukov et al (2018): high-level conditions under which sample splitting ensures $\hat{\eta} \xrightarrow{p} \eta$

- Tricky to verify these conditions
- Depends on (unknown) “sparsity” of DGP

Duration elasticities < 0.2

TABLE 1—DURATION ELASTICITIES FROM OBSERVATIONAL MTURK DATA

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
log reward	0.186	−0.0600					
	(0.0947)	(0.0585)					
log reward-ML res.			−0.0958	−0.0787	−0.198	−0.181	−0.0299
			(0.00558)	(0.00651)	(0.0281)	(0.0161)	(0.00402)
Observations	644,873	629,756	644,873	629,756	93,775	292,746	258,352
Clusters	41,167	26,050	41,167	26,050	6,962	18,340	24,923
Type	OLS	FE	ML	ML–FE	ML	ML	ML
Data	Pooled	Pooled	Pooled	Pooled	2017	2016–2017	2014–2016

Notes: This table presents η estimates using data scraped from MTurk. Units are HIT batches. Column 1 presents the unadjusted coefficient from a bivariate regression of log duration on log reward. Column 2 estimates the specification in equation (2). Column 3 presents estimates from an OLS regression of the residualized log duration on the residualized log reward, as in equation (5) averaged across the two sample splits. Column 4 adds the fixed effects in column 2 as further controls to column 3. Columns 5–7 present the double ML estimate from different scraped subsamples. Standard errors are clustered at the requester level.

Experiments

Retention experiments

- Hire workers for a translation tasks at a common wage
- Then ask if they want to do the task again at an experimentally manipulated wage
- Get retention probability elasticity

Recruitment experiments

- Offer to hire workers to perform a new task at manipulated wage
- Get recruitment probability elasticity

Retention elasticities centered around 0.1

Table 2—Offer Acceptance and Offered Rewards from Retention Experiments

	(1)	(2)	(3)	(4)
Panel A. Horton et al. (2011) probability of accepting offer				
Reward	0.127	0.140	0.0861	0.0973
	(0.0219)	(0.0241)	(0.0292)	(0.0333)
Observations	328	307	125	107
η	0.234	0.241	0.192	0.202
SE	0.0334	0.0364	0.0594	0.0664
Average reward	11.60	11.63	11.37	11.50
Sophisticated	No	No	Yes	Yes
Controls	No	Yes	No	Yes
Panel B. Dube et al. (2017) probability of accepting offer				
Reward	0.0267	0.0486	0.0764	0.0782
	(0.0171)	(0.0202)	(0.0348)	(0.0329)
Controls	No	Yes	No	Yes
Observations	5184	5017	1702	1618
η	0.052	0.077	0.118	0.114
SE	0.0333	0.0322	0.0534	0.0479
Average reward	9	9	9	9
Sophisticated	No	No	Yes	Yes

Notes: Coefficients from equation (6) from "retention" experiments, and calculated elasticities, assessed at the specification sample mean. Units are individual workers. Robust standard errors in parentheses.

Recruitment elasticities <0.1

TABLE 3—RECRUITMENT ELASTICITIES FROM THREE EXPERIMENTS

	(1)	(2)	(3)	(4)
Reward	0.00186	0.0451	0.0287	0.00744
	(0.00188)	(0.0587)	(0.0104)	(0.00385)
Observations	600	1,800	338	2,738
η	0.0497	0.0724	0.115	0.0610
SE	0.0503	0.0944	0.0417	0.0290
Average reward	83.33	4	10.04	22.13
Experiment	Spot diff.	Classify reviews	Brainstorming	Pooled

Notes: Coefficients from equation (6) estimated from "recruitment" experiments, and calculated elasticities, assessed at the experimental sample mean. Units are individual workers. The pooled specification includes experiment fixed effects, and is weighted by the inverse of the standard deviation of rewards within each experiment. Robust standard errors in parentheses.

Summary

Even in a thick labor market, various measures of labor supply to the firm appear inelastic in the short run

Requesters that are in a hurry should (and probably do) pay higher wages that are still below their private valuations

How different would the reward distribution be if requesters were required to be price takers?

- Would a minimum wage reduce efficiency here?
- What if there were a separate minimum wage for “urgent” projects?
- How would the rewards distribution change if employers bid on workers?

Staiger, Spetz, Phibbs (2010)

Employment prospects of nurses closely tied to local hospitals

Are RN wages suppressed below MPL?

Test for strategic dependence in wage setting (oligopsony)

- Nurse Pay Act of 1990: VA hospitals switch from national wage scale to matching local competitors
- Initial degree of under / over- payment provides an IV for VA wage
- See if non-VA hospitals respond or are price takers

VAs that underpaid experience large boost

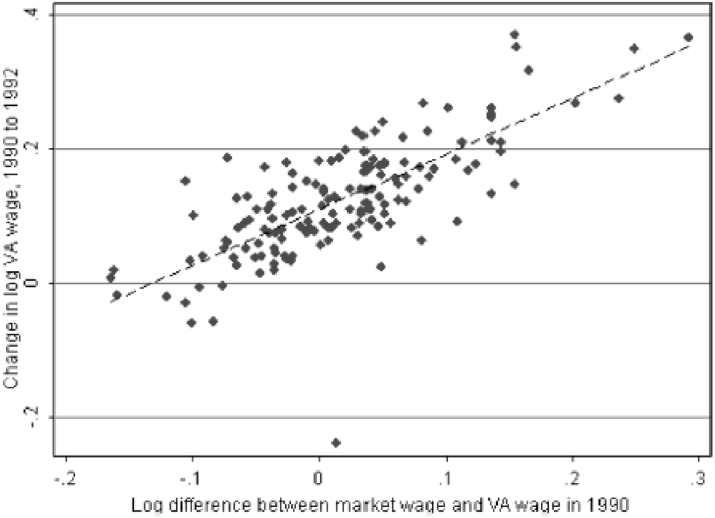


FIG. 1.—Difference between the market wage and the VA wage in 1990 and its association with the change in the VA wage from 1990 to 1992. Each point represents data for a single VA hospital in our sample, with the simple regression line for these data also displayed.

VA wage gap strongly predicts non-VA wage growth

Table 3
Reduced-Form Estimates of the Impact of the VA Wage Gap in 1990 on the Wage Changes in VA and Non-VA Hospitals, 1990–92

Independent Variables	VA Only (1)	Non-VA Only (2)	Non-VA Only (3)	Non-VA Only (4)	Non-VA Only (5)
Wage gap at nearest VA in 1990 (log(market wage) – log(VA wage))	.830	.090	.161	.345	.344
	(.055)	(.034)	(.061)	(.067)	(.065)
Wage gap at nearest VA in 1990 × dummy if > 15 miles to VA			−.109	−.154	−.146
			(.075)	(.072)	(.071)
Wage gap at nearest VA in 1990 × dummy if > 30 miles to VA			−.033	−.112	−.120
			(.064)	(.091)	(.091)
Dummy if > 15 miles to VA					−.008
					(.006)
Dummy if > 30 miles to VA					.000
					(.008)
MSA dummies?	No	No	No	Yes	Yes
R^2	.559	.011	.017	.281	.282
No. of observations	155	1,179	1,179	1,179	1,179

Implied LS to non-VA hospitals very low

Table 5
Two-Stage Least Squares Estimates of RN Labor Supply Elasticities

Independent Variables	(1)	(2)	(3)	(4)	(5)	(6)
Change in the log wage gap between hospital and its two nearest competitors	.076	.080	.016	.185	.185	.127
	(.137)	(.133)	(.177)	(.138)	(.135)	(.185)
Dummy if VA hospital			.023			.019
			(.014)			(.014)
MSA dummies?	No	No	Yes	No	No	Yes
"FAR" instruments included?	No	Yes	Yes	No	Yes	Yes
"GAP" instruments used?	No	No	No	Yes	Yes	Yes
p-value for test of the over-identifying restrictions	.71	.45	.31	.20	.20	.12
No. of observations	1,334	1,334	1,334	1,334	1,334	1,334

Summary

Strong evidence of strategic dependence in wage setting

Implied LS elasticity to hospitals ~ 0.1

- But are wages really set according to exploitation index?
- How to distinguish from "collusion"?

Ongoing work

- Spillovers from company-specific min wages (Derenoncourt, Noelke, Weil, 2020)
- Spillovers from actual min wage (Haanwinckel, 2018)
- Links between concentration and wage setting (Berger, Herkenhoff, Mongey, 2019; Arnold, 2019)

Kline, Petkova, Williams, Zidar (2019)

Study effect of winning a patent on firm productivity and wages using treasury tax files

Patents are designed to provide firms w/ temporary monopoly rights: are monopoly rents shared w/ workers?

- Patent grants a truly firm-specific shock
- Competitive benchmark: wages shouldn't adjust

1st time patenting firms are small (median firm size = 17)

- Unlikely to have much market power over new hires
- But potentially have power over incumbents

Main findings:

- Patents raise productivity
- And wages of incumbent workers
- But not entry wages

Obtaining a US patent (crash course)

Discover a novel, non-obvious, useful idea

Submit application to USPTO central office ("filing date")

- Central office routes application to the supervisory patent examiner (SPE) of the appropriate *art unit*
- SPE assigns application to a patent examiner

Examiner issues an initial decision ("initial decision date")

- Allowance (roughly 10% of initial decisions) or "rejection"
- "Rejection" is a revise & resubmit
- Applicant and examiner may engage in many rounds of revision

Research design

Two valuable patent applications submitted by two separate firms to the USPTO in the same year

Research design

Two valuable patent applications submitted by two separate firms to the USPTO in the same year

They are routed to the same art unit

Research design

Two valuable patent applications submitted by two separate firms to the USPTO in the same year

They are routed to the same art unit

One is initially allowed and the other is not

Research design

Two valuable patent applications submitted by two separate firms to the USPTO in the same year

They are routed to the same art unit

One is initially allowed and the other is not

Assume parallel trends for initially allowed/rejected patents (DiD)

- Validate w/ event studies + balance tests + low-value patents

Problem: Many patents worthless

Solution: predict ex-ante value using app characteristics

Problem: Many patents worthless

Solution: predict ex-ante value using app characteristics

Kogan, Papanikolaou, Seru, and Stoffman (2017; KPSS)

- Estimate excess stock return responses to patent grant announcements
- Empirical bayes posterior valuations ξ_j for each patent j

Problem: Many patents worthless

Solution: predict ex-ante value using app characteristics

Kogan, Papanikolaou, Seru, and Stoffman (2017; KPSS)

- Estimate excess stock return responses to patent grant announcements
- Empirical bayes posterior valuations ξ_j for each patent j

Use ξ_j to identify valuable patents in a broader sample

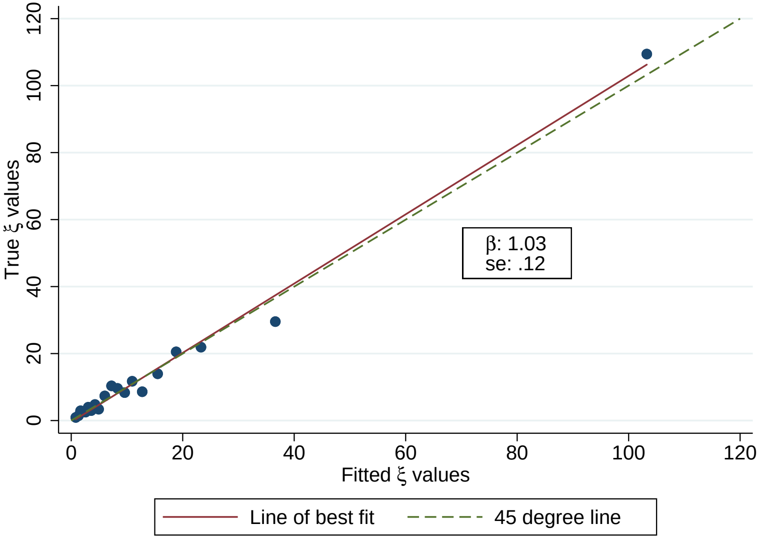
- Fit RE Poisson QML explaining ξ_j in terms of firm and application characteristics that are fixed at the time of application
- Extrapolate to non-public firms and to rejected applications
- Very strong explanatory power ($R^2 = .69$)

Poisson model

	KPSS value (ξ)	
$\mathbf{1}$(patent family size $= 1$)	0.28	(0.06)
log(patent family size)	0.23	(0.04)
$\mathbf{1}$(number of claims $= 1$)	0.68	(0.19)
log(number of claims)	0.30	(0.03)
$\mathbf{1}$(revenue $= 0$)	1.42	(0.14)
log(revenue)	0.14	(0.02)
$\mathbf{1}$(employees $= 0$)	0.45	(0.07)
log(employees)	-0.01	(0.02)
application year	-0.03	(0.05)
(application year)^2	-0.01	(0.01)
decision year	0.30	(0.06)
(decision year)^2	-0.03	(0.01)
constant	-1.40	(0.21)
log(σ)	0.24	(0.05)
N 596	# groups	260

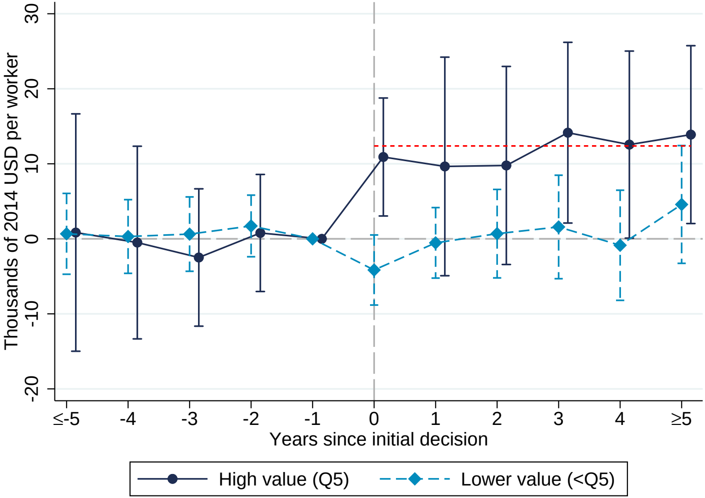
Notes: Random effects are by art unit. Standard errors are in parentheses.

Predicted vs. actual patent value



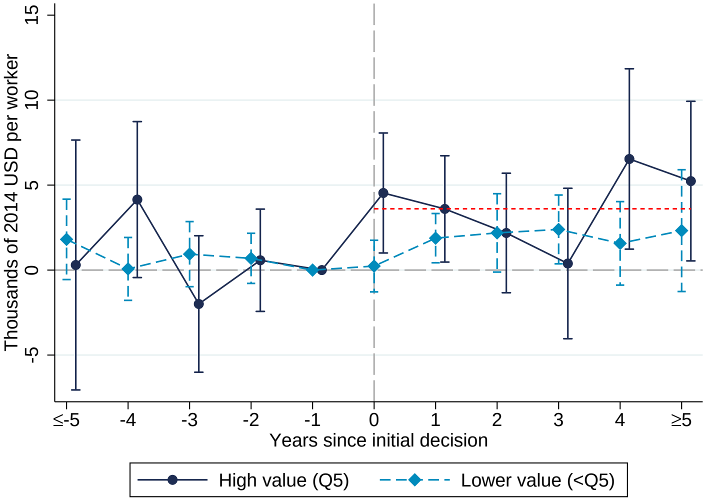
Notes: The fitted ξ values on the x-axis are obtained from a Poisson model of ξ on the DWPI count of unique countries where the application was filed, the number of claims in the application, the application year, the initial decision year, the revenue of the firm in the year of application, the number of employees in the application year, and art unit random effects.

Event study: Surplus (EBITD + W2) per worker



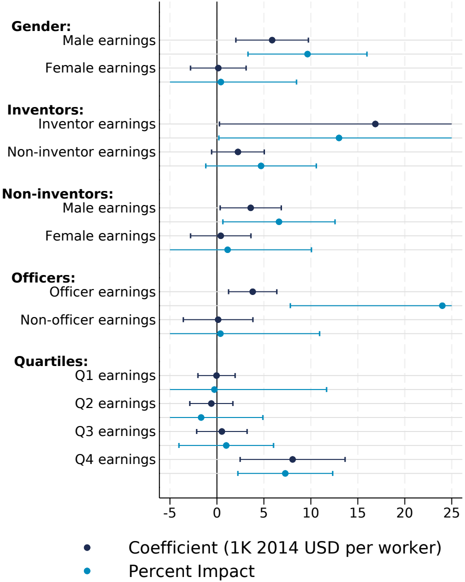
Notes: Two-way standard errors are clustered by (1) art unit, and (2) application year by decision year. Regressions include art unit by application year by calendar year fixed effects and firm fixed effects. Values along the x-axis for the $Q5$ series are offset from their integer value to improve readability. Surplus is EBITD (earnings before interest, tax, and depreciation) + W2 wage bill. $Q5$ is quintile 5 of predicted patent value. $< Q5$ are the remaining four quintiles. 95% confidence intervals shown. Dotted red line is pooled DID impact for a top quintile patent application receiving an initial allowance post-decision.

Event study: Wage bill per worker



Notes: Two-way standard errors are clustered by (1) art unit, and (2) application year by decision year. Regressions include art unit by application year by calendar year fixed effects and firm fixed effects. Values along the x-axis for the $Q5$ series are offset from their integer value to improve readability. $Q5$ is quintile 5 of predicted patent value. $< Q5$ are the remaining four quintiles. 95% confidence intervals shown. Dotted red line is pooled DID impact for a top quintile patent application receiving an initial allowance post-decision.

Within firm inequality



No impact on earnings of new hires..

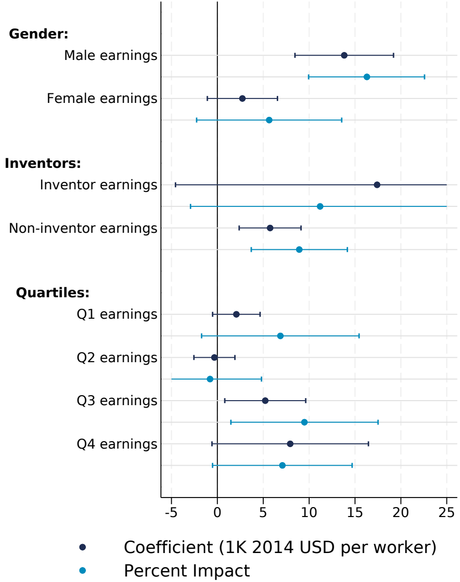
TABLE VII
EARNINGS IMPACTS BY YEAR OF ENTRY/EXIT

						Change since application year		
	Avg cohort earnings (1)	Avg stayer earnings (2)	Avg leaver earnings (3)	Avg entrant earnings (4)	Avg recent earnings (5)	Avg stayer earnings (6)	Avg leaver earnings (7)	Avg entrant earnings (8)
High value ($Q5$)	3.96	7.78	− 1.54	0.11	− 2.71	6.50	2.77	0.95
	(2.29)	(2.93)	(1.94)	(1.64)	(1.81)	(3.10)	(5.65)	(1.80)
Mean of outcome ($Q5$)	57.39	72.56	50.57	33.01	41.59	72.56	50.57	33.01
% impact ($Q5$)	6.9	10.7	− 3.0	0.3	− 6.5	9.0	5.5	2.9
Lower value ($<Q5$)	0.34	2.48	0.90	0.31	0.78	1.48	− 3.87	− 0.18
	(1.18)	(1.59)	(1.39)	(0.79)	(1.01)	(1.63)	(2.40)	(0.70)
Observations	151,892	99,558	109,169	70,079	68,691	99,558	109,169	70,079

Notes. This table reports difference-in-differences estimates of the effect of initial patent allowances on worker outcomes for employees who stay, enter, and exit, separately for high and low ex ante valuable patent applications. Estimates correspond to coefficients on interactions of the designated value category with a postdecision indicator and an indicator for the application being initially allowed. Controls include main effect of value category interacted with a postdecision indicator, firm fixed effects, and art unit by application year by calendar year fixed effects, as in equation (8). Standard errors (reported in parentheses) are two-way clustered by (i) art unit and (ii) application year by decision year. "Avg cohort earnings" measures the W2 earnings of workers employed by the firm in the year of application. "Avg stayer earnings" measures the W2 earnings of workers employed by the firm in the year of application who are also employed in the present year. "Avg leaver earnings" measures the W2 earnings of workers employed by the firm in the year of application who are not employed in the present year. "Avg entrant earnings" measures the W2 earnings of employees who were not employed by the firm in the previous year. "Avg recent entrant earnings" tracks the average earnings of employees hired by the firm within the past three years. "Change since application year" columns are earnings measures in the current year (columns (2), (3), and (4)) minus their respective values in the application year. Earnings are measured in thousands of 2014 dollars.

Impacts concentrated among firm stayers

Within-firm heterogeneity: Firm Stayers

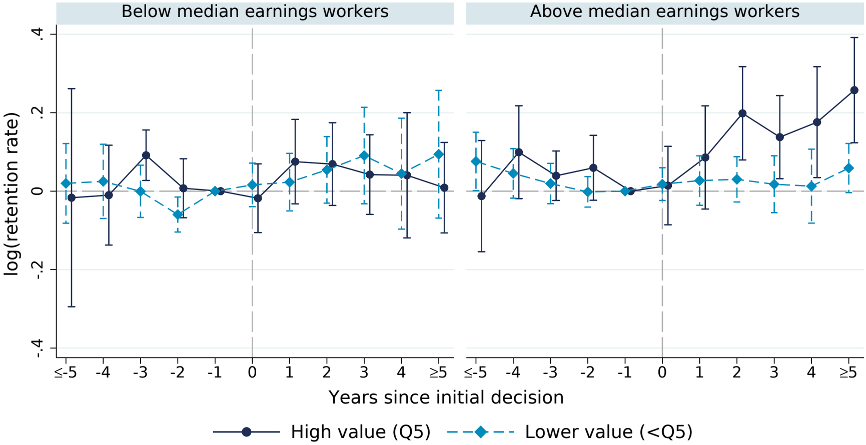


Instrumenting raises pass through estimates

TABLE VIII
PASS-THROUGH ESTIMATES

	Wage bill per worker		Avg male earnings		Avg noninventor earnings		Avg stayer earnings		Avg stayer earnings minus earnings in app yr		Avg noninventor stayer earnings	
	OLS (1)	IV (2)	OLS (3)	IV (4)	OLS (5)	IV (6)	OLS (7)	IV (8)	OLS (9)	IV (10)	OLS (11)	IV (12)
Panel A: Surplus/worker												
Impact	0.16	0.29	0.18	0.53	0.13	0.19	0.20	0.61	0.19	0.51	0.16	0.48
	(0.01)	(0.12)	(0.01)	(0.18)	(0.00)	(0.11)	(0.01)	(0.30)	(0.01)	(0.27)	(0.00)	(0.22)
Elasticity	0.19	0.35	0.19	0.54	0.17	0.25	0.19	0.56	0.18	0.47	0.17	0.50
Observations	103,437	103,437	95,004	95,004	100,901	100,901	99,558	99,558	99,558	99,558	94,909	94,909
First-stage ***F***		12.12		10.60		9.34		13.38		13.38		8.93
Exogeneity		.29		.08		.60		.14		.22		.08
Anderson-Rubin 90% CI		(0.10,0.57)		(0.27,0.98)		(−0.01,0.43)		(0.21,1.36)		(0.11,1.14)		(0.21,1.18)
Panel B: Value added/worker												
Impact	0.07	0.23	0.08	0.38	0.06	0.15	0.08	0.49	0.06	0.41	0.06	0.41
	(0.00)	(0.13)	(0.00)	(0.19)	(0.00)	(0.10)	(0.00)	(0.28)	(0.00)	(0.25)	(0.00)	(0.20)
Elasticity	0.15	0.47	0.14	0.67	0.13	0.33	0.13	0.78	0.10	0.65	0.11	0.72
Observations	103,437	103,437	95,004	95,004	100,901	100,901	99,558	99,558	99,558	99,558	94,909	94,909
First-stage ***F***		8.99		8.27		8.79		9.39		9.39		6.86
Exogeneity		.17		.08		.35		.08		.11		.03
Anderson-Rubin 90% CI		(0.07,0.61)		(0.15,1.03)		(−0.01,0.38)		(0.15,1.39)		(0.08,1.15)		(0.18,1.17)

Retention response concentrated among "top half"



Separation-wage elasticity of ~ 1.5

Retention of Application Cohort

	All (1)	Above median (2)	Men (3)	Women (4)	Noninventors (5)
Retention elasticity	1.22	1.41	0.80	1.17	1.31
	(0.58)	(0.65)	(0.35)	(0.80)	(0.68)
Separation elasticity	− 1.62	− 2.76	− 1.14	− 1.73	− 1.66
Observations	99,558	81,728	88,100	71,591	94,909
First-stage F	7.81	5.80	31.13	3.61	6.74
Exogeneity	.034	.029	.041	.060	.047
Anderson-Rubin 90% CI	(0.459, 3.080)	(0.597, 4.091)	(0.283, 1.524)	(0.233, 8.687)	(0.422, 3.655)

Notes. This table reports IV estimates of the effect of increases in selected earnings measures on the retention of employees. The excluded instrument is the interaction of the top quintile of ex ante value $\hat{\xi}$ category with a postdecision indicator and an indicator for the application being initially allowed. Controls include the main effect of value category interacted with a postdecision indicator and interaction of lower quintile value category with a postdecision indicator interacted with an indicator for initially allowed, firm fixed effects, and art unit by application year by calendar year fixed effects. Standard errors (reported in parentheses) are two-way clustered by (i) art unit and (ii) application year by decision year. "Separation Elasticity" is computed from the retention elasticity via a Taylor approximation. Specifically, the separation elasticity estimate is $-\frac{\bar{R}}{1-\bar{R}}\hat{\varepsilon}$, where $\hat{\varepsilon}$ is the IV estimate of the elasticity of retentions with respect to the wage and $\bar{R}$ is the mean retention rate among firms with high ex ante value patents. "Exogeneity" reports a p-value for the test of the null hypothesis that IV and OLS estimators have the same probability limit. "Above median" refers to members of the application cohort who earned above that firm's median in the application year. Stayers are defined as those who were employed by the same firm in the year of application. Earnings are measured in thousands of 2014 dollars.

Rent sharing redux

Firm shocks matter for worker wages, even when firms are small

But pass through is unequal across groups

- Men get more than women
- Incumbents more than new hires
- Inventors more than non-inventors

No one model to rule them all

- CCHK model would be misleading here – wage responses not proportional to hiring responses
- Important to separate retention and recruitment margins, especially when training / hiring costs substantial
- Pay for performance?

Comparable to economy-wide studies? (e.g., Garin and Silverio, 2017; Lamadon, Mogstad, Seltzer, 2019)

Jäger, Schoefer, Young, Zweimüller (2020)

Vanilla DMP model says wages (w) set via Nash bargaining to divide match surplus:

$$\underbrace{W + J}_{\text{value of match}} \quad - \quad \underbrace{(U + V)}_{\text{outside options}}$$

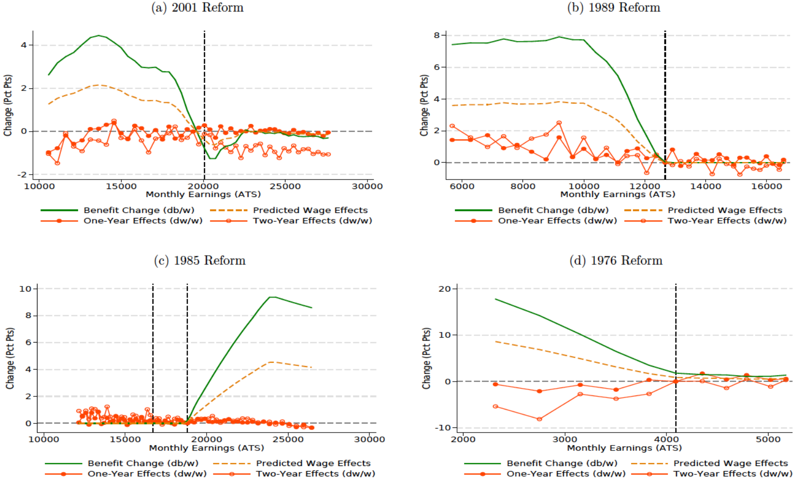
- Implies wages sensitive to value of unemployment U
- Under continuous renegotiation should apply to both incumbent workers and new hires

Test if wages sensitive to increase in UI generosity

- Diff-in-diff using Austrian reforms to UI benefit
- Key finding: no effect on incumbent wage growth

Little effect on wage growth of incumbents

Figure 4: Nonparametric Benefit Changes and Wage Effects



Note: The figure plots reform-induced replacement rate changes and wage effects for all four reform. Observations are binned by their base year (year before the reform was enacted) earnings percentile on the x-axis. The dashed orange line indicates the wage growth that the reform would induce in the calibrated bargaining model with a wage-benefit sensitivity of 0.48. The red circles indicate the wage effects that the reform induced at the one- and two-year horizon. Section 4.2 provides more information.

Mixed evidence on wage response of new hires

Table 4: Wage Effects by Individual Labor Market Status Transition Types

	Panel A: Effects by Transition Type							
	Full Sample		Job Stayers		Recalled Workers		Job Movers	
Time Horizon	1-Year	2-Year	1-Year	2-Year	1-Year	2-Year	1-Year	2-Year
Est. Wage Effect	-0.014	-0.022	-0.026	-0.027	0.061	0.007	0.054	-0.035
	(0.016)	(0.030)	(0.013)	(0.021)	(0.139)	(0.137)	(0.096)	(0.090)
Base-Year Transition Rate			0.828	0.705	0.040	0.057	0.069	0.110
Mincer + Ind.-Occ. FEs	X	X	X	X	X	X	X	X

	Panel B: Employment-Unemployment-Employment Movers							
	1-Year Earnings Effects				2-Year Earnings Effects			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Est. Wage Effect	-0.372	-0.215	-0.337	-0.249	-0.126	-0.104	-0.064	-0.115
	(0.146)	(0.140)	(0.161)	(0.228)	(0.147)	(0.154)	(0.194)	(0.237)
Base-Year Transition Rate	0.022	0.022	0.022	0.022	0.035	0.035	0.035	0.035
Transition-Specific Controls		X		X		X		X
Mincer + Ind.-Occ. FEs	X	X	X	X	X	X	X	X
Firm FEs			X	X			X	X

- Wrong signed effect on stayers and EUE movers
- But can't rule out positive effects on recalled workers / all job movers..

Di Addario, Kline, Saggio, Sølvsten (2020)

Large class of "sequential auction" models predict wages depend not just on current but also prior firm (Postel-Vinay and Robin, 2002; Bagger, Lentz, Postel-Vinay, Robin, 2016)

More general principle: outside options *at the time of hire* should affect the wage

Do firms price discriminate based on where the workers are hired from?

Examine using Italian wage records

- Records include the reason for each job separation (e.g., fired, laid off, resignation)
- Measure hiring wage as average earnings in 1st year on the job

Preliminaries: coding job transitions

Job histories of workers $i \in \{1, ..., n\}$ across job matches $m \in \{1, ..., M_i\}$.

- $Q_{im} = 1$ iff worker i quits match m ("EE transition")
- *Destination* firm is $j(i, m) \in \{1, ..., J\}$

Origin firm/state is

$$h(i, m) = \begin{cases} j(i, m-1), & \text{if } Q_{i,m-1} = 1 \text{ and } m > 1, \\ U, & \text{if } Q_{i,m-1} = 0 \text{ and } m > 1, \\ N, & \text{if } m = 1, \end{cases}$$

- U is "hired from non-employment"
- N is "new labor force entrant."

Dual Wage Ladder (DWL) specification

The log *hiring* wage for worker i in match m is:

$$y_{im} = \underbrace{\alpha_i}_{\text{worker effect}} + \underbrace{\psi_{j(i,m)}}_{\text{destination effect}} + \underbrace{\lambda_{h(i,m)}}_{\text{origin effect}} + X'_{im}\delta + \varepsilon_{im}.$$

- Similar to AKM model for *mean* wage in a match + "origin effect" for firm/state from which worker was hired
- O/D effs capture "where you're from" vs "where you're at"

Dual Wage Ladder (DWL) specification

The log *hiring* wage for worker i in match m is:

$$y_{im} = \underbrace{\alpha_i}_{\text{worker effect}} + \underbrace{\psi_{j(i,m)}}_{\text{destination effect}} + \underbrace{\lambda_{h(i,m)}}_{\text{origin effect}} + X'_{im}\delta + \varepsilon_{im}.$$

- Similar to AKM model for *mean* wage in a match + "origin effect" for firm/state from which worker was hired
- O/D effs capture "where you're from" vs "where you're at"

Treat $\{\alpha_i\}_{i=1}^{N}, \{\psi_j, \lambda_j\}_{j=1}^{J}$ as unrestricted fixed effects

- Note: each firm is a separate 2D type!
- SA models traditionally restrict $\psi_j = \psi(p_j)$, $\lambda_j = \lambda(p_j)$ [PVR, 2002a,b; Cahuc et al, 2006; Bagger et al, 2016; Bagger and Lentz, 2019]

Exogenous mobility

Let $\varepsilon_i = (\varepsilon_{i1}, ..., \varepsilon_{iM_i})'$ and $\mathcal{W}_i = \{j(i,m), h(i,m), X_{im}, \alpha_i\}_{m=1}^{M_i}$

We assume

$$\mathbb{E}[\varepsilon_i | \mathcal{W}_i] = 0.$$

- Rules out selection on time-varying component present at time of hiring.
- Does *not* prohibit selection on (ψ, λ)
- Implied by standard SA models, which typically assume efficient mobility along stable job-ladder in p

Dynamics: three examples

Career Path #1: two EUE transitions ($Q_{i1} = 0, Q_{i2} = 0$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)}$$

Dynamics: three examples

Career Path #1: two EUE transitions ($Q_{i1} = 0, Q_{i2} = 0$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)}$$

Career path #2: two EE transitions ($Q_{i1} = 1, Q_{i2} = 1$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{j(i,1)}$$

Dynamics: three examples

Career Path #1: two EUE transitions ($Q_{i1} = 0, Q_{i2} = 0$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)}$$

Career path #2: two EE transitions ($Q_{i1} = 1, Q_{i2} = 1$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_{j(i,1)}$$

Career path #3: EUE followed by EE ($Q_{i1} = 0, Q_{i2} = 1$)

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)} + \lambda_{j(i,2)} - \lambda_U$$

Observations:

- Path #1 yields destination based wage growth ala AKM
- Path #2 vs #3: wage penalty of $\lambda_{j(i,1)} - \lambda_U$ for displacement

The PVR model

PVR show that the poaching wage ϕ must satisfy:

$$U\left(\phi\left(\epsilon, p, q\right)\right) = U\left(\epsilon q\right) - \kappa \int_{q}^{p} \bar{F}\left(x\right) U'\left(\epsilon x\right) dx$$

where $\bar{F}\left(x\right) = 1 - F\left(x\right)$ and $\kappa = \frac{\lambda_1}{\rho+\delta+\mu}$ is fn of offer arrival, discount rate, etc.

The PVR model

PVR show that the poaching wage ϕ must satisfy:

$$U\left(\phi\left(\epsilon, p, q\right)\right) = U\left(\epsilon q\right) - \kappa \int_{q}^{p} \bar{F}\left(x\right) U'\left(\epsilon x\right) dx$$

where $\bar{F}\left(x\right) = 1 - F\left(x\right)$ and $\kappa = \frac{\lambda_1}{\rho+\delta+\mu}$ is fn of offer arrival, discount rate, etc.

If $U\left(x\right) = \ln x$ then poaching wage can be written:

$$\ln \phi\left(\epsilon, p, q\right) = \underbrace{\ln \epsilon}_{\text{worker type}} + \underbrace{\ln q}_{\text{poached firm type}} - \underbrace{\kappa \int_{q}^{p} \frac{\bar{F}(x)}{x} dx}_{\text{option val of type upgrade}}$$

- Poaching wage is decreasing in the productivity gap between poaching and poached firms (compensating diff)

DWL representation

By Fund Thm of Calculus, option value can be written

$$\kappa \int_q^p \frac{\bar{F}(x)}{x} dx = I(q) - I(p), \text{ where}$$

$$I(z) \equiv \kappa \int_z^\infty \frac{\bar{F}(x)}{x} dx \text{ is upgrade from } z \text{ to } p_{max}$$

DWL representation

By Fund Thm of Calculus, option value can be written

$$\kappa \int_q^p \frac{\bar{F}(x)}{x} dx = I(q) - I(p), \text{ where}$$

$$I(z) \equiv \kappa \int_z^\infty \frac{\bar{F}(x)}{x} dx \text{ is upgrade from } z \text{ to } p_{max}$$

Implies poaching wages obey log-linear reduced form:

$$\ln \phi(\varepsilon, p, q) = \underbrace{\ln \varepsilon}_{=\alpha(\varepsilon)} + \underbrace{I(p)}_{=\psi(p)} + \underbrace{\ln q - I(q)}_{=\lambda(q)}$$

- $\psi'(p) < 0$ (comp diff for expected wage growth)
- $\lambda'(q) > 0$ (tougher to poach from more productive firm)
- Exogenous mobility: worker goes to more productive firm

Properties of O/D effs

$$\ln \phi(\varepsilon, p, q) = \underbrace{\ln \varepsilon}_{=\alpha(\varepsilon)} + \underbrace{l(p)}_{=\psi(p)} + \underbrace{\ln q - l(q)}_{=\lambda(q)}$$

1. Productivity identified from sum of firm's O+D effs:

$$\psi(p) + \lambda(p) = \ln p$$

2. O/D effs are negatively correlated across firms:

$$\mathbb{C}(\psi(p), \lambda(p)) < 0$$

3. Excess variance of O vs D effs:

$$\mathbb{V}[\lambda(p)] > \mathbb{V}[\psi(p)]$$

Bagger et al (2014) extension

BF-PVR allow workers to extract a share $\beta \in [0,1]$ of rent.

Bagger et al (2014) extension

BF-PVR allow workers to extract a share $\beta \in [0,1]$ of rent.

Optimal poaching wage becomes:

$$\begin{aligned} \ln \phi\left(\epsilon, p, q, \mathcal{X}, \mathcal{E} \mid \beta\right) &= \alpha(\epsilon) + g(\mathcal{X}) + \mathcal{E} \\ &+ \underbrace{\beta \ln p + I(p \mid \beta)}_{=\psi(p)} + \underbrace{(1-\beta)\ln q - I(q \mid \beta)}_{=\lambda(q)}, \end{aligned}$$

where $\mathcal{X}$ is labor market experience, $\mathcal{E}$ is a transitory shock to worker productivity, and $I(z \mid \beta) = (1-\beta)^2 \kappa \int_z^\infty \frac{\bar{F}(x)/x}{1+\kappa\beta\bar{F}(x)} dx$ is decreasing in z and β.

Bagger et al (2014) extension

BF-PVR allow workers to extract a share $\beta \in [0,1]$ of rent.

Optimal poaching wage becomes:

$$\begin{aligned} \ln \phi(\epsilon, p, q, \mathcal{X}, \mathcal{E} \mid \beta) &= \alpha(\epsilon) + g(\mathcal{X}) + \mathcal{E} \\ &+ \underbrace{\beta \ln p + I(p \mid \beta)}_{=\psi(p)} + \underbrace{(1-\beta)\ln q - I(q \mid \beta)}_{=\lambda(q)}, \end{aligned}$$

where $\mathcal{X}$ is labor market experience, $\mathcal{E}$ is a transitory shock to worker productivity, and $I(z \mid \beta) = (1-\beta)^2 \kappa \int_z^\infty \frac{\bar{F}(x)/x}{1+\kappa\beta\bar{F}(x)} dx$ is decreasing in z and β.

Observe that:

- As $\beta \to 0$, BF-PVR$\to$PVR
- As $\beta \to 1$, BF-PVR$\to$AKM! (no origin effs)

O/D effs in BF-PVR

$$\begin{aligned} \ln \phi(\epsilon, p, q, \mathcal{X}, \mathcal{E} \mid \beta) &= \alpha(\epsilon) + g(\mathcal{X}) + \mathcal{E} \\ &+ \underbrace{\beta \ln p + I(p \mid \beta)}_{=\psi(p)} + \underbrace{(1-\beta)\ln q - I(q \mid \beta)}_{=\lambda(q)} \end{aligned}$$

- Productivity identified by $\psi(p) + \lambda(p) = \ln p$
- But large β can overcome comp. diff:

$$\beta > 1/2 \Rightarrow \psi'(p) > 0 \Rightarrow \mathbb{C}(\psi(p), \lambda(p)) > 0$$

- Shape restrictions
  1. Origin effs *concave* in $\ln p$: $\frac{d^2}{d(\ln p)^2}\lambda(p) < 0$
  2. Dest effs *convex* in $\ln p$: $\frac{d^2}{d(\ln p)^2}\psi(p) > 0$

Bounds on worker bargaining power

Consider *firm*-level variance components (firm-size weighted):

$$\mathbb{V}_J[\psi], \quad \mathbb{V}_J[\lambda], \quad \mathbb{C}_J[\psi, \lambda],$$

Bounds on worker bargaining power

Consider *firm*-level variance components (firm-size weighted):

$$\mathbb{V}_J[\psi], \quad \mathbb{V}_J[\lambda], \quad \mathbb{C}_J[\psi, \lambda],$$

- Excess variance of destination effects places lower bound on bargaining strength:

$$\beta \geq \frac{1}{2} + \frac{\mathbb{V}_J[\psi] - \mathbb{V}_J[\lambda]}{2\mathbb{V}_J[\psi + \lambda]}.$$

Intuition: as β grows, we approach AKM specification

Bounds on worker bargaining power

Consider *firm*-level variance components (firm-size weighted):

$$\mathbb{V}_J[\psi], \quad \mathbb{V}_J[\lambda], \quad \mathbb{C}_J[\psi, \lambda],$$

- Excess variance of destination effects places lower bound on bargaining strength:

  $$\beta \geq \frac{1}{2} + \frac{\mathbb{V}_J[\psi] - \mathbb{V}_J[\lambda]}{2\mathbb{V}_J[\psi + \lambda]}.$$

  Intuition: as β grows, we approach AKM specification

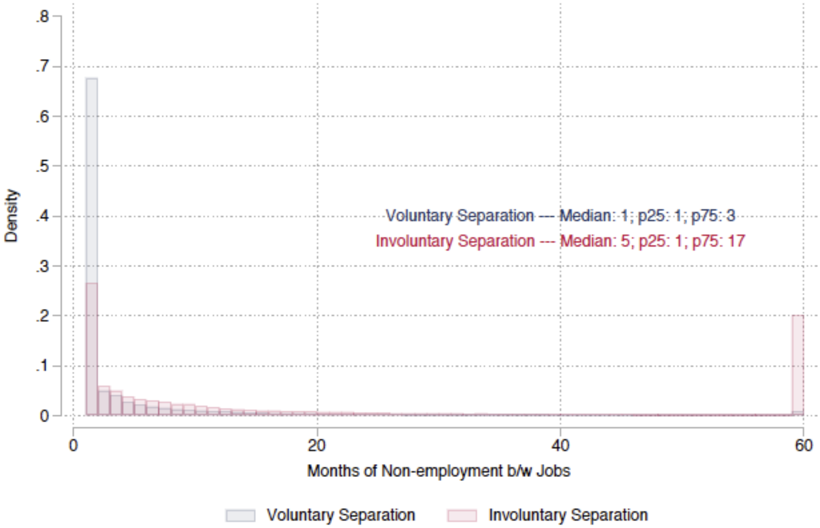
- $\beta > 1/2 \Rightarrow$inequality restriction on O/D eff correlation:

  $$\rho_J(\psi, \lambda) \geq \sqrt{\frac{\mathbb{V}_J[\psi]}{\mathbb{V}_J[\psi + \lambda]}} \left(1 - \frac{3}{10}\sqrt{\frac{\mathbb{V}_J[\lambda]}{\mathbb{V}_J[\psi + \lambda]}}\right)$$

  Intuition: $\beta > 1/2 \Rightarrow$ O/D effs both increasing in p

Median resignation yields job next month

Median time between jobs for other separations 5 months



Diagnostic #1: Is there a wage penalty for displacement?

Two workers i and ℓ transition between the same firms j and k

- Worker i has EE ("voluntary") transition

$$\mathbb{E}[y_{i2} - y_{i1} \mid \mathcal{W}_i] = \psi_k - \psi_j + \lambda_j - \lambda_N$$

- Worker ℓ has EUE ("involuntary") transition

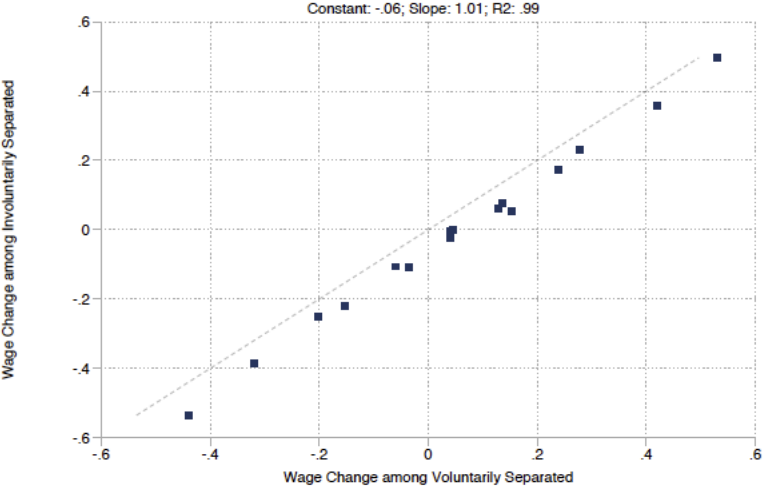
$$\mathbb{E}[y_{\ell 2} - y_{\ell 1} \mid \mathcal{W}_\ell] = \psi_k - \psi_j + \lambda_U - \lambda_N$$

Penalty for involuntary separation is

$$\begin{aligned} \lambda_j - \lambda_U &= \mathbb{E}[y_{i2} - y_{i1} \mid \mathcal{W}_i] \\ &- \mathbb{E}[y_{\ell 2} - y_{\ell 1} \mid \mathcal{W}_\ell] \end{aligned}$$

Rather than exact match on first two employers, group workers by coworker wage quartile at jobs #1 & #2 (16 groups)

Roughly constant penalty



Note: Each dot represents the adjusted log hiring wage change from job#1 to job#2 for different combinations of origin/destination quartiles of mean-coworkers wages. These dots are computed for two groups of workers. The first group (x-axis) corresponds to workers that voluntarily quit their first job. The second group (y-axis) corresponds to workers that were involuntarily separated from their first job.

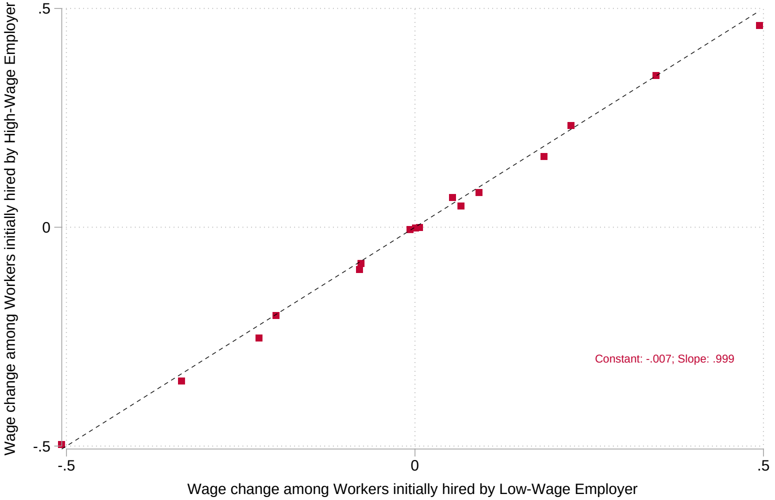
Diagnostic #2: Does it matter who lays you off?

Recall that DWL model predicts making 2 involuntary transitions ($Q_{i1} = 0, Q_{i2} = 0$) yields AKM style model of wage changes:

$$\mathbb{E}[y_{i3} - y_{i2} \mid \mathcal{W}_i] = \psi_{j(i,3)} - \psi_{j(i,2)}$$

- Identity $j(i, 1)$ of first employer is excludable!
- Test by comparing workers whose first employer was in top / bottom tercile of coworker wages

1st job irrelevant for workers displaced twice



Roughly 4% penalty for hiring from non-employment

(Note: we have normalized $\lambda_N = 0$)

Table 4: Variance Decomposition across Person-Job Observations --- DWL Model

	Pooled	**Men**	**Women**
Std Dev of log hiring wages	0.5286	0.4706	0.5623
Mean origin effect among involuntarily separated	0.0556	0.0536	0.0687
Mean origin effect among voluntarily separated	0.0561	0.0543	0.0690
Origin effect when hired from unemployment (λ_U)	0.0163	0.0136	0.0220
Bias-Corrected Variance Components			
Std Dev of worker effects	0.2823	0.2479	0.2798
Std Dev of destination firm effects	0.2580	0.2434	0.2828
Std Dev of origin effects	0.0439	0.0454	0.0431
Std Dev of origin effects (among poached workers)	0.0761	0.0782	0.0798
Correlation of worker, destination firm effects	0.3157	0.2351	0.3441
Correlation of worker, origin effects	0.1200	0.1629	0.0757
Correlation of destination firm, origin effects	0.0316	0.0308	0.0000
Percent of Total Variance Explained by			
Worker effects	28.52%	27.75%	24.77%
Destination firm effects	23.81%	26.74%	25.29%
Origin effects	0.69%	0.93%	0.59%
Covariance of worker, destination	16.46%	12.81%	17.23%
Covariance of worker, origin	1.06%	1.66%	0.58%
Covariance of destination, origin	0.26%	0.31%	0.00%
X'δ and associated covariances	1.66%	3.51%	0.09%
Residual	27.55%	26.30%	31.46%

It ain't where you're from..

Table 4: Variance Decomposition across Person-Job Observations --- DWL Model

	Pooled	Men	Women
Std Dev of log hiring wages	0.5286	0.4706	0.5623
Mean origin effect among involuntarily separated	0.0556	0.0536	0.0687
Mean origin effect among voluntarily separated	0.0561	0.0543	0.0690
Origin effect when hired from unemployment (λ_U)	0.0163	0.0136	0.0220
Bias-Corrected Variance Components			
Std Dev of worker effects	0.2823	0.2479	0.2798
Std Dev of destination firm effects	0.2580	0.2434	0.2828
Std Dev of origin effects	0.0439	0.0454	0.0431
Std Dev of origin effects (among poached workers)	0.0761	0.0782	0.0798
Correlation of worker, destination firm effects	0.3157	0.2351	0.3441
Correlation of worker, origin effects	0.1200	0.1629	0.0757
Correlation of destination firm, origin effects	0.0316	0.0308	0.0000
Percent of Total Variance Explained by			
Worker effects	28.52%	27.75%	24.77%
Destination firm effects	23.81%	26.74%	25.29%
Origin effects	0.69%	0.93%	0.59%
Covariance of worker, destination	16.46%	12.81%	17.23%
Covariance of worker, origin	1.06%	1.66%	0.58%
Covariance of destination, origin	0.26%	0.31%	0.00%
X'δ and associated covariances	1.66%	3.51%	0.09%
Residual	27.55%	26.30%	31.46%

Note: This table reports the variance decomposition based upon the DWL model across person-job observations. We also report the (firm-size weighted) corresponding average of the origin effects for individuals that were involuntary separated as well as the estimated origin effect when hired from unemployment. All origin effects are

Dest effs $\approx 14\times$ as variable as orig effs across firms

Table 5: Variance Decomposition across Firms

	Pooled	Men	Women
# of firms with identified destination and origin effect	297,865	201,080	99,508
Bias-Corrected Variance Components			
Std of Destination Effects	0.2590	0.2449	0.2724
Std of Origin Effects	0.0707	0.0721	0.0510
Correlation of destination, origin	0.2511	0.2491	0.3168
Std of Destination + Origin Effects	0.2851	0.2720	0.2926
Lower Bound on Bargaining Power	0.8819	0.8703	0.9182
Lower Bound on Correlation of Destination, Origin Effects	0.8409	0.8288	0.8824

Note: Here we report the variance decomposition across firms where each firm has an identified origin and destination firm effect. Variance components are weighted by average firm-size over 2005-2015 as recorded by official INPS records collected in the dataset *Anagrafica* , see text for details. Variance components corrected using the leave-out bias correction of Kline, Saggio and Sølvsten (2020). The lower bounds on the barganing power and correlation of destination and origin firm effects are based upon equation (5)-(6), see text for details.

Implied std dev of log productivity=.28

Compare to std log VA/L≈0.8

Table 5: Variance Decomposition across Firms

	Pooled	**Men**	**Women**
# of firms with identified destination and origin effect	297,865	201,080	99,508
Bias-Corrected Variance Components			
Std of Destination Effects	0.2590	0.2449	0.2724
Std of Origin Effects	0.0707	0.0721	0.0510
Correlation of destination, origin	0.2511	0.2491	0.3168
Std of Destination + Origin Effects	$\boxed{0.2851}$	0.2720	0.2926
Lower Bound on Bargaining Power	0.8819	0.8703	0.9182
Lower Bound on Correlation of Destination, Origin Effects	0.8409	0.8288	0.8824

Note: Here we report the variance decomposition across firms where each firm has an identified origin and destination firm effect. Variance components are weighted by average firm-size over 2005-2015 as recorded by official INPS records collected in the dataset *Anagrafica* , see text for details. Variance components corrected using the leave-out bias correction of Kline, Saggio and Sølvsten (2020). The lower bounds on the barganing power and correlation of destination and origin firm effects are based upon equation (5)-(6), see text for details.

Need $\beta > .88$ to explain excess orig eff var

Which would require O/D corr $> .84$, but empirical corr is only .25..

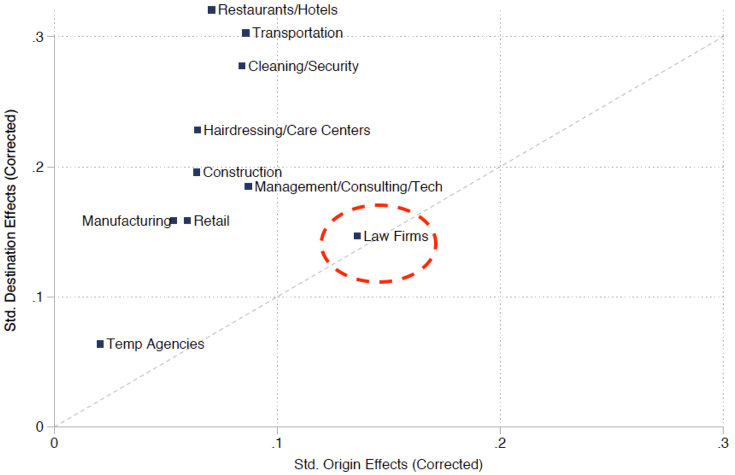
Table 5: Variance Decomposition across Firms

	Pooled	**Men**	**Women**
# of firms with identified destination and origin effect	297,865	201,080	99,508
Bias-Corrected Variance Components			
Std of Destination Effects	0.2590	0.2449	0.2724
Std of Origin Effects	0.0707	0.0721	0.0510
Correlation of destination, origin	0.2511	0.2491	0.3168
Std of Destination + Origin Effects	0.2851	0.2720	0.2926
Lower Bound on Bargaining Power	0.8819	0.8703	0.9182
Lower Bound on Correlation of Destination, Origin Effects	0.8409	0.8288	0.8824

Note: Here we report the variance decomposition across firms where each firm has an identified origin and destination firm effect. Variance components are weighted by average firm-size over 2005-2015 as recorded by official INPS records collected in the dataset *Anagrafica* , see text for details. Variance components corrected using the leave-out bias correction of Kline, Saggio and Sølvsten (2020). The lower bounds on the barganing power and correlation of destination and origin firm effects are based upon equation (5)-(6), see text for details.

Heterogeneity: law firms have important origin effs

Figure 4: Variability of Origin and Destination Effects by Sector



Note: This figure reports leave-out corrected standard deviations of destination and origin firm effects for selected sectors of the Italian economy (2-Digit 2007 Ateco codes). All variance components are firm-size weighted. The dashed line is the 45 degree line.

But even among law firms O/D correlation too low

Table 6: Variability of Origin and Destination Effects by Sector

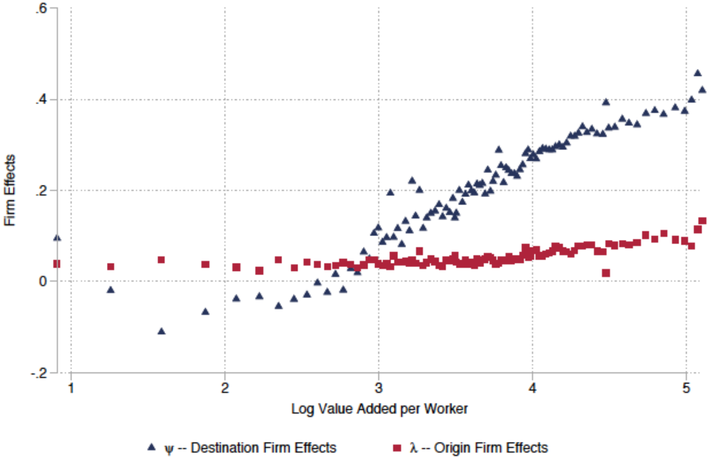
	SD of Destination Effects	SD of Origin Effects	Correlation of Origin, Destination Effects	Lower Bound on Bargaining Power	Lower Bound on Correlation
Retail	0.1585	0.0597	0.2260	0.8269	0.7868
Construction	0.1959	0.0639	-0.0693	0.9211	0.8786
Restaurants / Hotels	0.3206	0.0706	0.0675	0.9413	0.9018
Hairdressing / Care Centers	0.2284	0.0641	0.1399	0.8979	0.8567
Law Firms	0.1468	0.1359	0.0399	0.5369	0.5758
Manufacturing	0.1585	0.0536	0.2737	0.8409	0.7992
Transportation	0.3028	0.0859	-0.0632	0.9401	0.8969
Cleaning / Security	0.2777	0.0842	0.0874	0.8966	0.8551
Temp Agencies	0.0639	0.0206	0.1651	0.8702	0.8291
Management / Consulting / Tech	0.1847	0.0870	0.4190	0.7406	0.6991

Note: This table reports leave-out corrected standard deviations of destination and origin firm effects within selected sectors of the Italian economy (2-Digit 2007 Ateco codes). All variance components are firm-size weighted. The lower bounds on the barganing power and correlation of destination and origin firm effects are based upon equation (5)-(6), see text for details.

O/D effs both increasing in VA

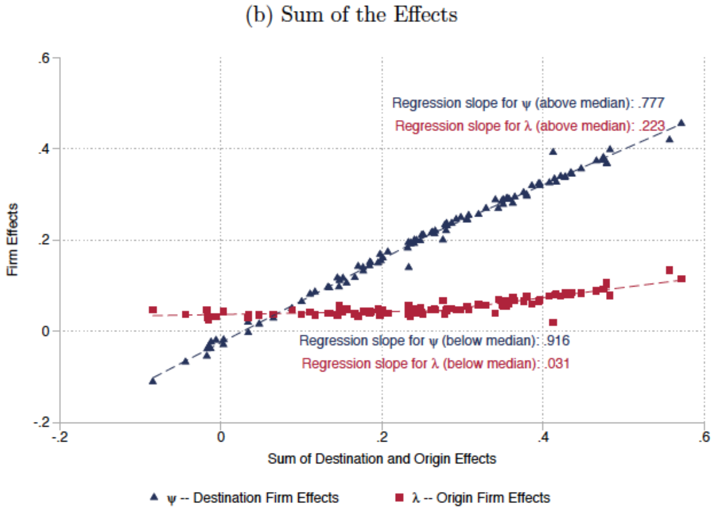
Figure 5: Origin and Destination Effects by Value Added

(a) Value Added per Worker



But violate shape restrictions

Also: BF-PVR requires $\beta > \max_{p'} d\psi(p') / d \ln p \approx 0.92$!

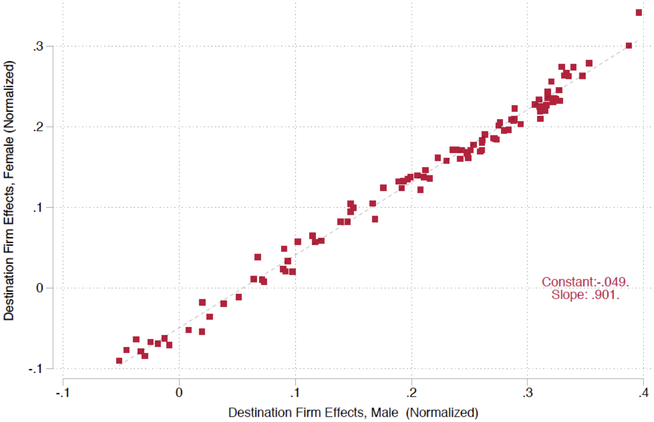


Note: each dot is mean within a VA bin (same as previous fig)

Female dest effs less sensitive to VA

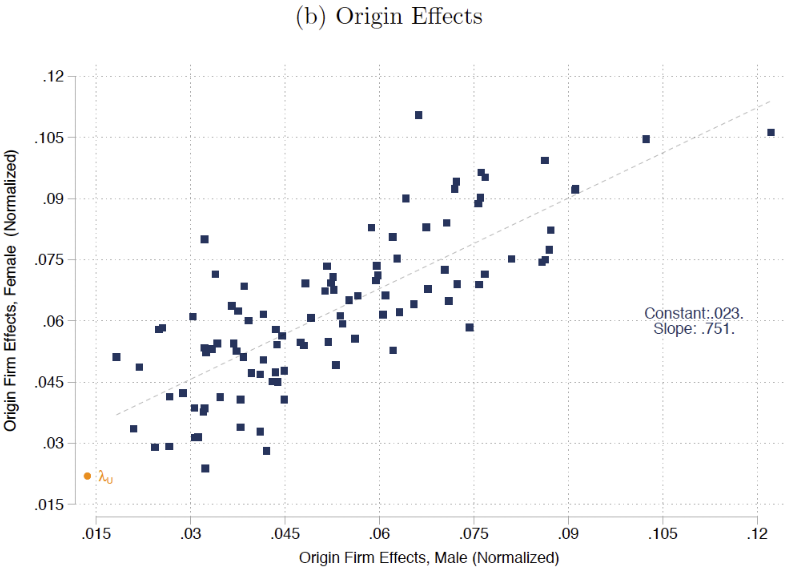
Figure 7: Origin and Destination Effects by Gender and Value Added

(a) Destination Effects



Same slope as found in Portugal [Card, Cardoso, Kline, 2015]

Same for orig effs but female suffer greater penalty for EUE

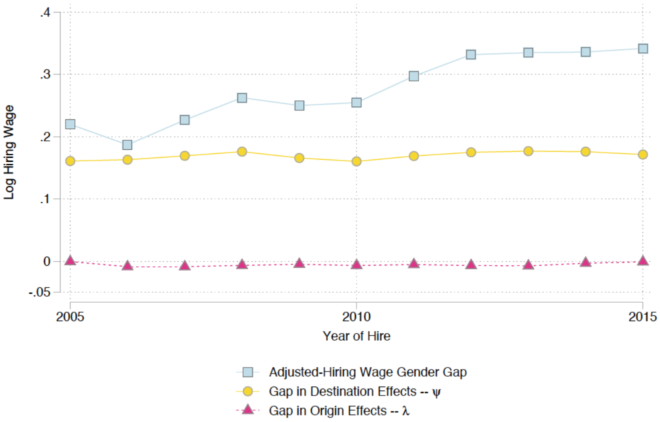


Where you're from irrelevant for gender gap

Initially explained by where you're at. Evolution due to other factors.

Figure 8: Gender Wage Gap and the DWL Model

(a) Entered Labor Market in 2005



Summary

Where you're hired from doesn't seem to matter quantitatively for most workers

Two notable exceptions:

- There is an important penalty for being hired from non-employment
- Highly skilled hierarchical professions (e.g. law) seem to exhibit origin effects

Why aren't hiring origins more important?

They likely are important for elite workers (NBA players, C-suite executives, star lawyers) who are expected to negotiate and typically have objective performance metrics that can be used to justify their pay

But most jobs commit to posted wages, likely for a mix of information and horizontal equity reasons

- Hall and Krueger (2012): bargaining only common among high skilled jobs in US
- Caldwell & Harmon (2019): in Denmark only 31% of manual and 51% of professional jobs engage in negotiation
- Postel-Vinay and Robin (2004): less productive firms commit not to match to avoid costly moral hazard $\Rightarrow$ dual labor markets
- Card, Moretti, Mas, and Saez (2012): horizontal inequity in pay generates potentially costly morale problems

References

1. Alfaro-Urena, A., Manelici, I., & Vásquez-Carvajal, J. P. (2019). The Effects of Multinationals on Workers: Evidence from Costa Rica.
2. Arnold, D. (2019). Mergers and acquisitions, local labor market concentration, and worker outcomes. Local Labor Market Concentration, and Worker Outcomes.
3. Azar, J., Berry, S., & Marinescu, I. (2019). Estimating Labor Market Power.
4. Bagger, J., Fontaine, F., Postel-Vinay, F., & Robin, J. M. (2014). Tenure, experience, human capital, and wages: A tractable equilibrium search model of wage dynamics. American Economic Review, 104(6), 1551-96.
5. Berger, D. W., Herkenhoff, K. F., & Mongey, S. (2019). Labor Market Power (No. w25719). National Bureau of Economic Research.
6. Berry, S. T. (1994). Estimating discrete-choice models of product differentiation. The RAND Journal of Economics, 242-262.
7. Berry, S., Levinsohn, J., & Pakes, A. (1995). Automobile prices in market equilibrium. Econometrica: Journal of the Econometric Society, 841-890.
8. Breiman, L. (2001). Random forests. Machine learning, 45(1), 5-32.

9. Caldwell, S., & Danieli, O. (2018). Outside options in the labor market.
10. Caldwell, S., & Harmon, N. (2019). Outside options, bargaining, and wages: Evidence from coworker networks. Unpublished manuscript, Univ. Copenhagen.
11. Card, D., Mas, A., Moretti, E., & Saez, E. (2012). Inequality at work: The effect of peer salaries on job satisfaction. American Economic Review, 102(6), 2981-3003.
12. Card, D., Cardoso, A. R., Heining, J., & Kline, P. (2018). Firms and labor market inequality: Evidence and some theory. Journal of Labor Economics, 36(S1), S13-S70.
13. Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., & Robins, J. (2018). Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal, 21(1), C1-C68.
14. Derenoncourt, E., Noelke, C., & Weil, D. 2020. Spillovers from voluntary employer minimum wages.
15. Dube, A., Jacobs, J., Naidu, S., & Suri, S. (2020). Monopsony in online labor markets. American Economic Review: Insights, 2(1), 33-46.
16. Di Addario, S., Kline, P., Saggio, R., & Solvsten, M. (2020). It ain't where you're from, it's where you're at: hiring origins, firm heterogeneity, and wages. Tech. Rep. 104-20, IRLE Working Paper.

17. Garin, A., & Silverio, F. (2017). How Does Firm Performance Affect Wages? Evidence from Idiosyncratic Export Shocks.
18. Haanwinckel, D. (2018). Supply, Demand, Institutions, and Firms: A Theory of Labor Market Sorting and the Wage Distribution.
19. Hall, R. E., & Krueger, A. B. (2012). Evidence on the incidence of wage posting, wage bargaining, and on-the-job search. American Economic Journal: Macroeconomics, 4(4), 56-67.
20. Jäger, S., Schoefer, B., Young, S., and Zweimüller, J. (2020). "Wages and the value of nonemployment." Quarterly Journal of Economics 135 (4).
21. Kogan, L., Papanikolaou, D., Seru, A., & Stoffman, N. (2017). Technological innovation, resource allocation, and growth. The Quarterly Journal of Economics, 132(2), 665-712.
22. Kline, P., Petkova, N., Williams, H., & Zidar, O. (2019). Who profits from patents? rent-sharing at innovative firms. The Quarterly Journal of Economics, 134(3), 1343-1404.
23. Lamadon, T., Mogstad, M., & Setzler, B. (2019). Imperfect Competition, Compensating Differentials and Rent Sharing in the US Labor Market (No. w25954). National Bureau of Economic Research.
24. Lindner, A., Muraközy, B., Reizer, B., & Schreiner, R. (2019). Technological Change and Skill Demand in Non-Competitive Labor Markets. working paper.

25. Manning, A., & Petrongolo, B. (2017). How local are labor markets? Evidence from a spatial job search model. American Economic Review, 107(10), 2877-2907.
26. Manning, A. (2020). Monopsony in labor markets: a review. International Labour Review.
27. Nimczik, J. S. (2017). Job mobility networks and endogenous labor markets.
28. Postel–Vinay, F., & Robin, J. M. (2002). Equilibrium wage dispersion with worker and employer heterogeneity. Econometrica, 70(6), 2295-2350.
29. Postel-Vinay, F., & Robin, J. M. (2004). To match or not to match?: Optimal wage policy with endogenous worker search intensity. Review of Economic Dynamics, 7(2), 297-330.
30. Rose, N. L. (1987). Labor rent sharing and regulation: Evidence from the trucking industry. Journal of Political Economy, 95(6), 1146-1178.
31. Staiger, D. O., Spetz, J., & Phibbs, C. S. (2010). Is there monopsony in the labor market? Evidence from a natural experiment. Journal of Labor Economics, 28(2), 211-236.