Fiscal Policy and Inequality

12. Fixed Effects Regression

Elliott Ash

15 October 2018

Empirical Application: Estimating Incidence

- **Evans, Ringel & Stech (1999)**: What is the incidence of cigarette taxes?
- How do changes in the cigarette tax affect prices?

Empirical Application: Estimating Incidence

- **Evans, Ringel & Stech (1999)**: What is the incidence of cigarette taxes?
- How do changes in the cigarette tax affect prices?
  - Does the burden fall on cigarette companies or smokers?
    - Why?
  - What are the welfare implications of this policy?

Evans, Ringel & Stech (1999)

- Cigarettes taxed at both federal and state level in US
  - Total revenue $\simeq$ \$35 billion per year, similar to estate taxation
- Variation in excise tax among states within US
  - from \$0.30 per pack in Virginia to \$4.35 in New York

Evans, Ringel & Stech (1999)

- Cigarettes taxed at both federal and state level in US
  - Total revenue $\simeq$ \$35 billion per year, similar to estate taxation
- Variation in excise tax among states within US
  - from \$0.30 per pack in Virginia to \$4.35 in New York
- Since 1975, more than 200 changes in state taxes
- Exploit these state-level changes using simple **diff-in-diff** research design

First Difference Estimator

- Basic idea: cigarette taxes raised in state A, but **not** in state B
  - Assume partial equilibrium (ignore other markets)

First Difference Estimator

- Basic idea: cigarette taxes raised in state A, but **not** in state B
  - Assume partial equilibrium (ignore other markets)
- Effect on prices in state A?

$$D = [P_{A1} - P_{A0}]$$

  - $D =$ "First difference" estimator

First Difference Estimator

- Basic idea: cigarette taxes raised in state A, but **not** in state B
  - Assume partial equilibrium (ignore other markets)
- Effect on prices in state A?

$$D = [P_{A1} - P_{A0}]$$

  - $D =$ "First difference" estimator
- Identification assumption (D): absent the tax change, cigarette prices in state A would not have changed between period 0 and 1.

Diff-in-Diff Estimator

- ID assumption is likely violated:
  - Prices fluctuate for many reasons (weather, demand...)
  - Hence, first difference estimator (D) likely to be biased

Diff-in-Diff Estimator

- ID assumption is likely violated:
  - Prices fluctuate for many reasons (weather, demand...)
  - Hence, first difference estimator (D) likely to be biased
- Better estimator: **diff-in-diff** (DD)

$$DD = [P_{A1} - P_{A0}] - [P_{B1} - P_{B0}]$$

- <u>Identification assumption</u> (DD): **"parallel trends"**
  - Absent tax change, trend in prices would have been the same in states A and B.

Diff-in-Diff Estimator

- ID assumption is likely violated:
  - Prices fluctuate for many reasons (weather, demand...)
  - Hence, first difference estimator (D) likely to be biased
- Better estimator: **diff-in-diff** (DD)

$$DD = [P_{A1} - P_{A0}] - [P_{B1} - P_{B0}]$$

- Identification assumption (DD): **"parallel trends"**
  - Absent tax change, trend in prices would have been the same in states A and B.
  - Allows for time-invariant differences between the two groups.

Diff-in-Diff Regression

- Can estimate the diff-in-diff effect using

$$P_{jt} = \alpha + \gamma \text{Treat}_{jt} + \lambda \text{After}_{jt} + \rho \text{Treat*After}_{jt} + \varepsilon_{jt}$$

  where Treat is a dummy for being the reform state, and After is a dummy for years after the reform.

Diff-in-Diff Regression

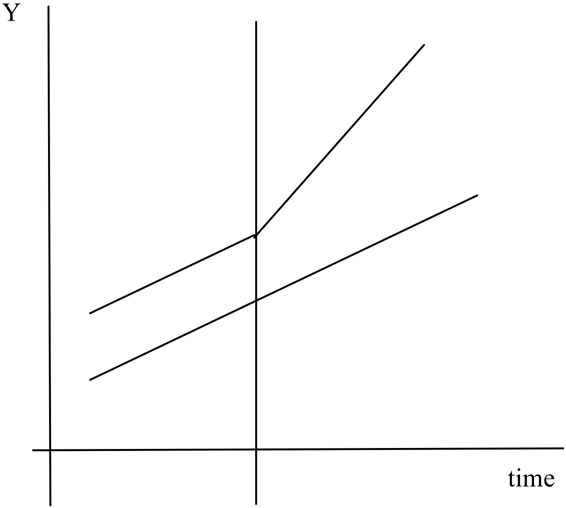
- Can estimate the diff-in-diff effect using

$$P_{jt} = \alpha + \gamma \text{Treat}_{jt} + \lambda \text{After}_{jt} + \rho \text{Treat*After}_{jt} + \varepsilon_{jt}$$

  where Treat is a dummy for being the reform state, and After is a dummy for years after the reform.

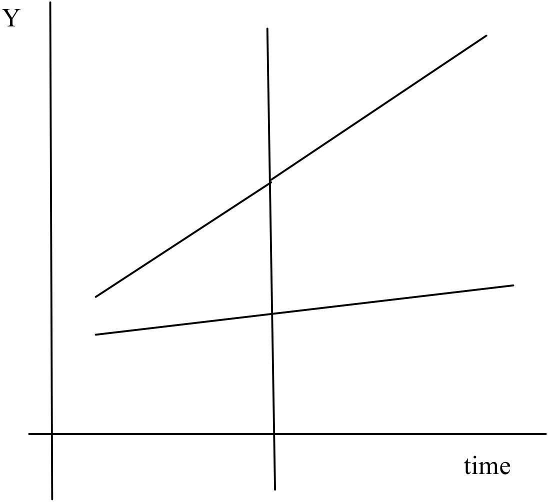
- Interpreting coefficients:
  - α, average in non-treated group, pre-treatment
  - γ, difference between treated and non-treated in pre-treatment period
  - λ, change in the control group after reform
  - ρ, the diff-in-diff treatment effect estimate (change in treatment group, relative to change in control group).

Diff-in-diff: Parallel trends assumption

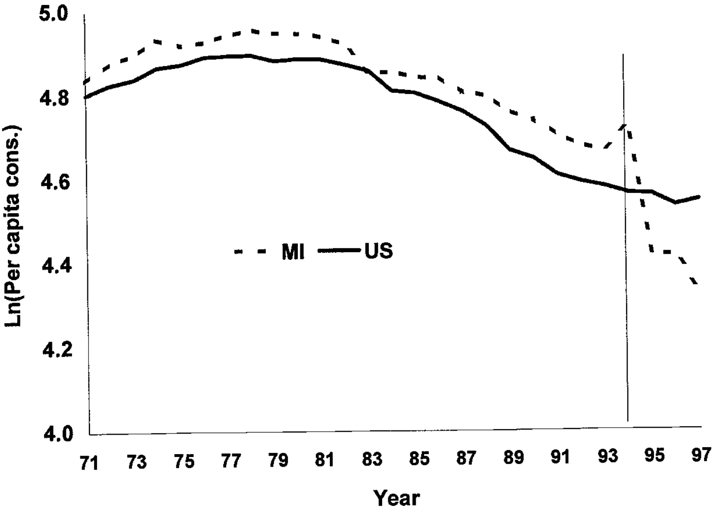


Diff-in-diff: Parallel trends assumption

Things don't always work as we wished they did...

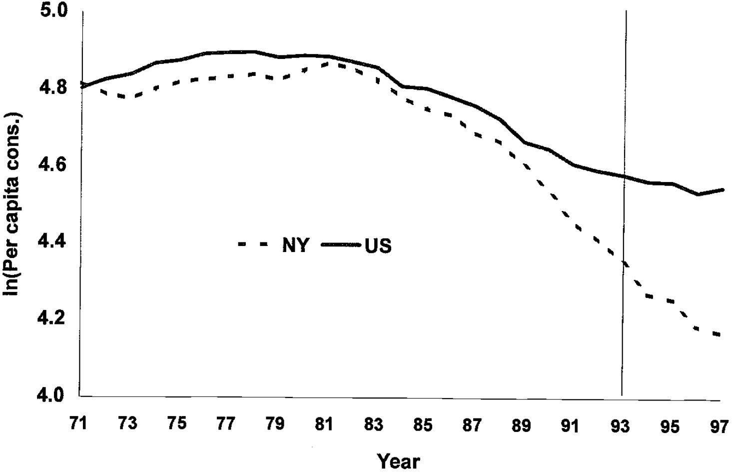


Evans, Ringel & Stech (1999)



Source: Evans, Ringel and Stech (1999)

Evans, Ringel & Stech (1999)



Source: Evans, Ringel and Stech (1999)

Evans, Ringel & Stech (1999)

- We can use “placebo” *DD* to test parallel trends assumption
  - Replicate *DD* estimate at other points in time when there was *no tax change*

Evans, Ringel & Stech (1999)

- We can use "placebo" *DD* to test parallel trends assumption
  - Replicate *DD* estimate at other points in time when there was *no tax change*
- If *DD* in other periods is not zero, then $DD_{t=1}$ likely biased
  - Useful to plot long time series of outcomes for treatment and control
  - Pattern should be parallel lines, with sharp change just after reform

Evans, Ringel & Stech (1999)

- In practice, Evans et al (1999) analyse many tax changes over 30 years (and 50 states)
- Useful to use **panel data** to obtain a <u>single incidence estimate</u>

Evans, Ringel & Stech (1999)

- In practice, Evans et al (1999) analyse many tax changes over 30 years (and 50 states)
- Useful to use **panel data** to obtain a single incidence estimate
- State j in year t applies policy (tax) T_{jt}

Evans, Ringel & Stech (1999)

- In practice, Evans et al (1999) analyse many tax changes over 30 years (and 50 states)
- Useful to use **panel data** to obtain a <u>single incidence estimate</u>
- State j in year t applies policy (tax) T_{jt}
- Want to estimate effect of T on price P. Cross-sectional OLS regression:

$$P_{jt} = \alpha + \beta T_{jt} + \varepsilon_{jt}$$

Evans, Ringel & Stech (1999)

- In practice, Evans et al (1999) analyse many tax changes over 30 years (and 50 states)
- Useful to use **panel data** to obtain a single incidence estimate
- State j in year t applies policy (tax) T_{jt}
- Want to estimate effect of T on price P. Cross-sectional OLS regression:

$$P_{jt} = \alpha + \beta T_{jt} + \varepsilon_{jt}$$

  - Problem: what if states with higher taxes T_{jt} also have more anti-tobacco campaigns?
    - $\text{Cov}\left(T_{jt}, \varepsilon_{jt}\right) \neq 0$ and $\hat{\beta}^{OLS}$ is **biased**.

Fixed-Effects Estimation

- Fixed-effects estimation:

$$P_{jt} = \alpha + \beta T_{jt} + \delta_j + \gamma_t + \varepsilon_{jt}$$

  - δ_j are state fixed effects
    - a dummy variable equaling one for state j's observations, and zero otherwise
  - γ_t are year fixed effects
    - a dummy variable equaling one for year t's observations, and zero otherwise
- Fixed-effects estimation generalises DD to $J > 2$ groups and $S > 2$ periods
- Requires panel (longitudinal) data

Fixed-Effects Estimation (FE)

- ID assumption (FE): absent the tax change, the trend in the outcome (price) would have been the same in treatment and control groups
  - In other words, there are no time-varying omitted variables.
  - Equivalent to "parallel trends"

Fixed-Effects Estimation (FE)

- ID assumption (FE): absent the tax change, the trend in the outcome (price) would have been the same in treatment and control groups
  - In other words, there are no time-varying omitted variables.
  - Equivalent to "parallel trends"
- Implicit assumption in FE: treatment effect is additive and constant (in this case, across states)

Fixed-Effects Estimation (FE)

- ID assumption (FE): absent the tax change, the trend in the outcome (price) would have been the same in treatment and control groups
  - In other words, there are no time-varying omitted variables.
  - Equivalent to "parallel trends"
- Implicit assumption in FE: treatment effect is additive and constant (in this case, across states)
- FE obtains identification from within-state variation over time
  - nationwide factors, such as federal tax changes, do not affect FE estimates

Threats to validity for FE and DiD

- If the groups are different in levels, maybe they evolve differently?
- Why did the treatment group adopt the policy, and not the control group?
- Policies are usually implemented in bundles (the timing of the treatment may not be by chance) $\rightarrow$ the outcome variable may be affected by these other policies
- The treatment should not affect the control group
- The composition of the treatment and control groups should not change as a result of treatment

Usual checks

- The two groups evolved similarly in the past (although this is not a guarantee for validity)
- The timing of the adoption of the policy was as good as random
- No other policies were adopted at the same time
- Verify that there is no reason to believe that the control group might be affected
- Add group-specific trends in outcome variable as additional regressor.

Let's Try It

```
# set index to group and time categories.
df.set_index(['state', 'year'], inplace=True)

# fit fixed effects model
from linearmodels.panel.PanelOLS import from_formula
model = from_formula('log_gdp ~ log_deathtax + EntityEffects + TimeEffects',
                     data=df)
results = model.fit()
results
```

What about all the othere regression outputs?

- Most are not that useful.
- R^2 tells you how much of the variance in the outcome is explained by the right-hand-side variables.
  - can be used to decide between two models, e.g., whether to take the log of your outcome variable.
  - can be arbitrarily increased by adding more variables; this is not a reason to add more variables.

Effect of cig taxes on cig prices (Evans et al 1999)

- Main regression model:

$$P_{jt} = \alpha + \beta T_{jt} + \delta_j + \gamma_t + \varepsilon_{jt}$$

  - P_{jt} = average retail price per pack in state j, year t (in \$ cents)
  - T_{jt} = total per pack tax (state+federal) in state j, year t (in \$ cents)
  - δ_j = state fixed effects. Control for any time-invariant differences in prices across states (eg, age distribution).
  - γ_t = year fixed effects. Control for shocks to prices that are common to all states but may vary across years (eg, federal tax change).

Evans, Ringel & Stech (1999): Results

TABLE 2
OLS Estimates, Retail Price Model: Tobacco Institute Data

Independent variable	Average state retail price, 1985–1996 Nominal (1)	Real (2)	Net retail price in Tennessee, 1970–1994 Nominal (3)	Real (4)
Nominal/real tax	1.01 (0.04)	0.92 (0.04)		
Nominal/real wholesale price			1.07 (0.02)	0.86 (0.04)
R^2	0.972	0.933	0.989	0.963
Observations	612	612	25	25

Standard errors in parentheses. Real prices in 1997 cents/pack. Models in columns (1) and (2) control for state effects.

Source: Evans, Ringel and Stech (1999)

Evans, Ringel & Stech (1999): Results

- 100% pass-through implies $\varepsilon_D = 0$ or $\varepsilon_S = \infty$ at state level
  - Why? Try using a graph to understand this

Evans, Ringel & Stech (1999): Results

- 100% pass-through implies $\varepsilon_D = 0$ or $\varepsilon_S = \infty$ at state level
  - Why? Try using a graph to understand this
- Why not $\varepsilon_D = 0$?
  - Would imply zero effect of taxes on consumption, which we can rule out (next slide).

Evans, Ringel & Stech (1999): Results

- 100% pass-through implies $\varepsilon_D = 0$ or $\varepsilon_S = \infty$ at state level
  - Why? Try using a graph to understand this
- Why not $\varepsilon_D = 0$?
  - Would imply zero effect of taxes on consumption, which we can rule out (next slide).
- Consider $\varepsilon_S = \infty$:
  - Tobacco companies can easily send inventories from high-tax to low-tax states, affecting prices by reducing supply
  - Pass-through would be lower at national level

Evans, Ringel & Stech (1999): Demand Elasticity

- To estimate ε_D, they use this model:

$$\ln(Q_{jt}) = \beta T_{jt} + X_{jt}\alpha + \mu_{1j} + \mu_{2j} \cdot \text{Time}_t + \nu_t + \varepsilon_{jt}$$

  - $\ln(Q_{jt})$ = log per capita consumption, state j, year t
  - T_{jt} = state+federal tax per pack, in \$ cents
  - μ_{1j} = state fixed effects
  - $\mu_{2j} \cdot \text{Time}_t$= state-specific time trends
  - ν_t = year fixed effects

Evans, Ringel & Stech (1999): Demand Elasticity

- To estimate ε_D, they use this model:

$$\ln\left(Q_{jt}\right) = \beta T_{jt} + X_{jt}\alpha + \mu_{1j} + \mu_{2j} \cdot \text{Time}_t + \nu_t + \varepsilon_{jt}$$

  - Notice that: $\beta = \frac{d \ln(Q)}{dT} \approx \frac{\Delta Q / Q}{\Delta T}$

Evans, Ringel & Stech (1999): Demand Elasticity

- To estimate ε_D, they use this model:

$$\ln\left(Q_{jt}\right) = \beta T_{jt} + X_{jt}\alpha + \mu_{1j} + \mu_{2j} \cdot \mathrm{Time}_t + \nu_t + \varepsilon_{jt}$$

  - Notice that: $\beta = \frac{d\ln(Q)}{dT} \approx \frac{\Delta Q/Q}{\Delta T}$
  - Recall that: $\varepsilon_D = \frac{\Delta Q}{\Delta P}\frac{\bar{P}}{\bar{Q}}$, where $\bar{P}$ and $\bar{Q}$ are mean values

Evans, Ringel & Stech (1999): Demand Elasticity

- To estimate ε_D, they use this model:

$$\ln\left(Q_{jt}\right) = \beta T_{jt} + X_{jt}\alpha + \mu_{1j} + \mu_{2j} \cdot \mathrm{Time}_t + \nu_t + \varepsilon_{jt}$$

  - Notice that: $\beta = \frac{d \ln(Q)}{dT} \approx \frac{\Delta Q / Q}{\Delta T}$
  - Recall that: $\varepsilon_D = \frac{\Delta Q}{\Delta P} \frac{\bar{P}}{\bar{Q}}$, where $\bar{P}$ and $\bar{Q}$ are mean values
  - Since $\varepsilon_S \approx \infty$, then $\Delta P \approx \Delta T$

Evans, Ringel & Stech (1999): Demand Elasticity

- To estimate ε_D, they use this model:

$$\ln\left(Q_{jt}\right) = \beta T_{jt} + X_{jt}\alpha + \mu_{1j} + \mu_{2j} \cdot \text{Time}_t + \nu_t + \varepsilon_{jt}$$

  - Notice that: $\beta = \frac{d\ln(Q)}{dT} \approx \frac{\Delta Q/Q}{\Delta T}$
  - Recall that: $\varepsilon_D = \frac{\Delta Q}{\Delta P}\frac{\bar{P}}{\bar{Q}}$, where $\bar{P}$ and $\bar{Q}$ are mean values
  - Since $\varepsilon_S \approx \infty$, then $\Delta P \approx \Delta T$
  - Hence:

$$\begin{aligned} \hat{\varepsilon}^D &= \frac{\Delta Q}{\Delta T}\frac{\bar{P}}{\bar{Q}} \\ \Rightarrow \hat{\varepsilon}^D &= \hat{\beta} \cdot \bar{P} \end{aligned}$$

Evans, Ringel & Stech (1999): Demand Elasticity

TABLE 3
OLS Estimates, Log Per Capita Consumption Model, Tobacco Institute Data, 1985–1996

Independent variable	Coefficients (standard errors) on: Real tax (1)	Real tax (2)	Real tax (3)	Real price (4)	Real price (5)	Real price (6)
Current value	−0.254 (0.037)	−0.165 (0.040)	−0.173 (0.041)	−0.176 (0.027)	−0.176 (0.027)	−0.167 (0.029)
1-year lag		−0.215 (0.413)	−0.188 (0.047)		−0.027 (0.032)	−0.031 (0.032)
2-year lag			−0.061 (0.045)			−0.017 (0.033)
Price elasticity	−0.424 (0.062)	−0.635 (0.074)	−0.705 (0.090)	−0.294 (0.045)	−0.337 (0.058)	−0.359 (0.072)
R^2	0.975	0.977	0.977	0.975	0.975	0.976

Standard errors in parentheses. The dependent variable is the log per capita consumption. There are 512 observations in each model. The mean price is $1.75/pack, and the mean per capita consumption is 105. All models include year effects, state effects, state-specific time trends, and log per capita consumption, plus measures of the fraction of adults in three age, three education, and two race groups.

Source: Evans, Ringel and Stech (1999)

Evans, Ringel & Stech (1999): Demand Elasticity

- We have:
  - $\hat{\beta} = -0.254$
  - $\bar{P} = \$1.75$
- Demand model estimate implies: $\varepsilon_D = -0.42$
  - 10% increase in price induces a 4.2% reduction in consumption
- This is the **short-run** elasticity. In the paper, they also estimate long-run elasticities.

Practice Exam Question

- Consider the market for diamonds in a closed economy. You estimate the following regression:

$$P_{jt} = \alpha + \beta T_{jt} + \delta_j + \gamma_t + \varepsilon_{jt}$$

where P_{jt} is the consumer price of diamonds in province j and year t, and T_{jt} is the excise tax on diamonds.

1. You obtain a point estimate $\widehat{\beta} = 0$. Interpret what this coefficient estimate means. What do you learn about the elasticity of supply and/or demand for diamonds?
2. Given your answer to (a) and economic intuition, who bears the incidence of this tax? Use the tax incidence formula.

A note on standard errors

- Imagine we want to estimate effect of a pill on longevity, sample is 1000 individuals each from Zurich and Zug.

A note on standard errors

- Imagine we want to estimate effect of a pill on longevity, sample is 1000 individuals each from Zurich and Zug.
- Two proposals for assigning to treatment and control:
  1. Flip a coin once: tail, individuals from Zurich are treated, heads, individuals from Zug are treated
  2. Flip coin 2000 times, once for each individual: tails, assigned to treatment; heads, assigned to control

A note on standard errors

- Imagine we want to estimate effect of a pill on longevity, sample is 1000 individuals each from Zurich and Zug.
- Two proposals for assigning to treatment and control:
  1. Flip a coin once: tail, individuals from Zurich are treated, heads, individuals from Zug are treated
  2. Flip coin 2000 times, once for each individual: tails, assigned to treatment; heads, assigned to control
- Which of the proposals is more informative about the impact of the treatment? Why?
  - OLS standard errors are similar in both cases.

A note on standard errors

- Imagine we want to estimate effect of a pill on longevity, sample is 1000 individuals each from Zurich and Zug.
- Two proposals for assigning to treatment and control:
  1. Flip a coin once: tail, individuals from Zurich are treated, heads, individuals from Zug are treated
  2. Flip coin 2000 times, once for each individual: tails, assigned to treatment; heads, assigned to control
- Which of the proposals is more informative about the impact of the treatment? Why?
  - OLS standard errors are similar in both cases.
- What's wrong? Presence of a common random effect:
  - In proposal 1, there might be some common shock affecting all individuals in the treatment group or in the control group.

A note on standard errors

- Imagine we want to estimate effect of a pill on longevity, sample is 1000 individuals each from Zurich and Zug.
- Two proposals for assigning to treatment and control:
  1. Flip a coin once: tail, individuals from Zurich are treated, heads, individuals from Zug are treated
  2. Flip coin 2000 times, once for each individual: tails, assigned to treatment; heads, assigned to control
- Which of the proposals is more informative about the impact of the treatment? Why?
  - OLS standard errors are similar in both cases.
- What's wrong? Presence of a common random effect:
  - In proposal 1, there might be some common shock affecting all individuals in the treatment group or in the control group.
  - OLS standard errors assume that all observations are independent realizations. Standard errors have to be corrected to account for the presence of a common random effect.

What about fixed effects regressions

- Consider the case of cigarette taxes. We have 50 states, times 50 years, equals 2500 observations.
    - if i only included the 10 years before and after the reform, i would have 20 years, or 1000 observations, but its essentially the same information, although standard errors would be bigger.
    - We need to account for *serial correlation* within state.

Solution: Clustering Standard Errors

- Cluster standard errors:
  - statistically acknowledges how many independent sources of information there are in the data.
  - can cluster by state and by time, to allow for correlation in errors along both dimensions.

Solution: Clustering Standard Errors

- Cluster standard errors:
  - statistically acknowledges how many independent sources of information there are in the data.
  - can cluster by state and by time, to allow for correlation in errors along both dimensions.
- In python linearmodels:

```
# fit fixed effects model with clustered errors
results = model.fit(cov_type='clustered',
                                cluster_entity=True,
                                cluster_time=True)
```

How to cluster?

- In general, cluster at the level of your treatment variation:
  - state-level reforms $\rightarrow$ cluster by state
  - city-level reforms $\rightarrow$ cluster by city
  - etc.

How to cluster?

- In general, cluster at the level of your treatment variation:
  - state-level reforms $\rightarrow$ cluster by state
  - city-level reforms $\rightarrow$ cluster by city
  - etc.
- Clustering is important, and in general will dramatically affect statistical significance of results (Bertrand, Duflo, and Mullainathan QJE 2004).