Lecture 16:
Sufficient Statistics

Kunal Sangani

ECON 416-1

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples

- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Presentation advice

- Keep slides clean.
  - One line per bullet unless absolutely necessary.
  - Even better if you don't need sub-bullets.
  - Model yourself on other presentation slides, not teaching slides.
- Good notation is worth +10 IQ points. Make your equations shine.
- Rule of thumb: 2 minutes per slide.
- Practice! Many senior professors deliver paper across seminars with *identical* script.

Presentation advice: Exhibits

- Graphs are 10x more persuasive than tables. Use graphs to convey, tables to verify.
- Read Eric Zwick's "A Graph is Worth a Thousand Citations."
- Text, figures, and tables should be legible from the back of the room.

Presentation advice: No stone left unturned

- First commandment of applied research: Know thy data.
  - How is it collected? Precise variable definitions?
- Be purposeful in analysis choices (e.g., variables, specification).
  - Don't do X just because literature does X, unless the point is contrast with the literature.
  - Is the empirical variable the same as its model-relevant counterpart?
  - Be clear about identifying assumptions, threats to causality.
  - Checklist when using IV, arguing for shift-share identification, or estimating spillovers.

Presentation advice: Miscellaneous

- John Cochrane's tips on papers apply to talks ("Writing Tips for Ph.D. Students").
  - After motivating the question, your task is to get to the central result as fast as possible.
  - Cochrane rule: *"There should be nothing before the main result that a reader does not need to know in order to understand the main result."*
  - Do not need extensive data description (especially of well-known datasets), etc.

Presentation advice: Miscellaneous

- John Cochrane's tips on papers apply to talks ("Writing Tips for Ph.D. Students").
  - After motivating the question, your task is to get to the central result as fast as possible.
  - Cochrane rule: *"There should be nothing before the main result that a reader does not need to know in order to understand the main result."*
  - Do not need extensive data description (especially of well-known datasets), etc.
- Jesse Shapiro's "How to Give an Applied Micro Talk: Unauthoritative Notes."
  - Audience does not need to see your process.
  - E.g., first I downloaded data from BLS, then collapsed data by state, then divided by 2008 CPI, then removed outliers.
  - Audience does need to know what the data they're seeing are.
  - E.g., average wage by state and year, excluding outliers ($>$\$100/hr in 2008 USD).

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples
- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Structural vs. reduced-form approaches

- Two paradigms for policy evaluation and welfare analysis.
- Structural approach: Specify and estimate complete model of economic behavior.
  - Transparent assumptions.
  - Illuminates mechanisms.
  - Can evaluate any policy (including unobserved counterfactual policies).
- Reduced-form approach: Evaluate experiments or quasi-experiments.
  - Hard to identify primitive parameters given selection, simultaneity, omitted variables.
  - Exogenous variation isolates causal effects.
  - Credible evaluation of policy outcomes (though not welfare).

Structural vs. reduced-form approaches: Critiques

- Critiques of reduced form approach:
  - Rosenzweig & Wolpin (2000): Natural experiments e.g. for returns to schooling give different estimates, and one could produce all numbers by varying assumptions about bias, heterogeneity, SUTVA. $\Rightarrow$ Reduced-form methods not identifying stable parameters.
  - Heckman & Vytacil (2005): IVs estimate LATE for marginal compliers.
  - Deaton (2009): Randomization is not a substitute for good economics.
  - Inability to illuminate mechanisms threatens external validity.
- Critiques of structural approach:
  - Lack of transparency about what identifies parameters.
  - Bayesian priors and functional form assumptions do work.
  - External validity only as good as model specification.

Sufficient statistics: A middle ground

- Public economists (e.g., Feldstein, Saez, Chetty) develop middle ground approach.
- Idea: Predict effects of policy using high-level elasticities instead of deep primitives.
  - Estimate these elasticities with transparent, reduced-form methods.
  - Do not need to discriminate between multiple sets of primitives that yield elasticities.
  - Formulas for welfare effects, optimal taxes, incidence, etc.

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples

- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Deadweight loss from a tax (Harberger 1964, Chetty 2009)

- Chetty (2009) uses deadweight loss from tax (Harberger 1964) as example.
- Suppose we have a consumer endowed with L units of numeraire good.
- Consumer has preferences over goods $j = 1, ..., J$ and quasilinear in consumption of numeraire good, y,
$$u(x_1, ..., x_J) + y.$$
- Producing x_j units of j'th good requires $c(x_j)$ units of numeraire good.
- Representative firm that takes prices as given and chooses quantities to solve:
$$\max_{\{x_j\}} \sum_j p_j x_j - c_j(x_j).$$

Deadweight loss from a tax (Harberger 1964, Chetty 2009)

- What is the efficiency loss from a tax t on good 1?
- Let $p(t)$ denote market-clearing prices as a function of tax rate.
- Consumer problem:

$$\max_x u(x) + y \qquad \text{s.t.} \qquad p(t) \cdot x + tx_1 + y = Z + T.$$

- Firm problem, using $c(x) = \sum_j c_j(x_j)$:

$$\max_x p(t) \cdot x - c(x).$$

- Social welfare:

$$W(t) = \underbrace{\max_x u(x) + (Z + T - p(t) \cdot x - tx_1)}_{\text{Money-metric consumer utility}} + \underbrace{\max_x p(t) \cdot x - c(x)}_{\text{Firm profits}}$$

Deadweight loss from a tax (Harberger 1964, Chetty 2009)

$$W(t) = \underbrace{\max_x u(x) + (Z + T - p(t)\cdot x - tx_1)}_{\text{Money-metric consumer utility}} + \underbrace{\max_x p(t)\cdot x - c(x)}_{\text{Firm profits}}$$

- Taking derivative with respect to t (using Envelope theorem):

$$\begin{aligned}\frac{dW(t)}{dt} &= \frac{dT}{dt} - \frac{dp(t)}{dt}\cdot x - x_1 + \frac{dp(t)}{dt}\cdot x \\ &= \frac{d\left(tx_1\right)}{dt} - x_1 \\ &= t\frac{dx_1}{dt}.\end{aligned}$$

- All we need to know is how equilibrium quantity of taxed good changes with t.
- Do not need to know full supply and demand system!

Deadweight loss from a tax (Harberger 1964, Chetty 2009)

- Sufficient statistic for welfare effect of tax change from t_1 to t_2:

$$\Delta W = \int_{t_1}^{t_2} t \frac{dx_1}{dt}(t) dt.$$

- Note that dx/dt is an *equilibrium* elasticity:

$$\frac{dx_1}{dt} = \frac{\partial x_1}{\partial p_1}\frac{\partial p_1}{\partial t} + \frac{\partial x_2}{\partial p_2}\frac{\partial p_2}{\partial t} + \ldots + \frac{\partial x_J}{\partial p_J}\frac{\partial p_J}{\partial t}.$$

- Sufficient statistics approach: Measure t and dx_1/dt to estimate welfare effect of tax.

- Structural alternative: Specify model of consumer demand and firm production, and solve for each individual component.

Deadweight loss from a tax (Harberger 1964, Chetty 2009)

- While sufficient statistics approach appealing in this context, some trade-offs.
- Derivation did not allow for pre-existing distortions in other markets.
  - Tax change could interact with other distortions (Theory of Second Best).
  - E.g., Goulder and Williams (2003): *"We show that under typical conditions the simple "excess-burden triangle" formula substantially underestimates the excess burden of commodity taxes, in some cases by a factor of 10 or more. This formula performs poorly because it ignores general equilibrium interactions—most important, interactions between the taxed commodity and the labor market."*
  - Would need a different sufficient statistics formula to handle these cases.
- How do we get dx_1/dt?
  - Moreover, may be interested in evaluating a large new tax out of sample support.
  - Extrapolating dx_1/dt to larger tax implicitly invoking a log-linear (or some other) approximation of demand?

Sufficient statistics: Generality

- Nevertheless, sufficient statistics approach overcomes the “curse of dimensionality” that accompanies modeling realistic heterogeneity.

- Example 1: suppose we now have N individuals, with income Z^i and preferences

$$u^i(x^i) + y.$$

Now need to estimate $\approx N \times J$ parameters $\partial x^i_j / \partial p_j$ and J parameters $\partial p_j / \partial t$.

- Utilitarian social welfare is:

$$W(t) = \sum_{i=1}^{N} \max_{x^i} \left(u^i(x^i) + Z_i + T_i - p(t) \cdot x_i - t x_1^i \right) + \max_{x} p(t) \cdot x - c(x).$$

$$\Rightarrow \frac{dW(t)}{dt} = \sum_{i=1}^{N} \frac{dT_i}{dt} - x_1^i = t \frac{d \sum_{i=1}^{N} x_1^i}{dt} = t \frac{dx_1}{dt}.$$

Sufficient statistics: Generality

- Example 2: Suppose now that each good j produced by different firm.
- Quantities x and welfare W will generally depend on firm conduct—whether firms maximize individual profits or whether they collude and maximize joint profits:

$$W(t) = \max_x u(x) + (Z + T - p(t) \cdot x - tx_1) + \sum_j \max_{x_j} p_j(t) x_j - c_j(x). \qquad \text{(Indiv.)}$$

versus

$$W(t) = \max_x u(x) + (Z + T - p(t) \cdot x - tx_1) + \max_x p(t) \cdot x - c(x). \qquad \text{(Collusion)}$$

- Need to estimate not just production technologies but also conduct.
- But in both cases,

$$\frac{dW(t)}{dt} = t \frac{dx_1}{dt}.$$

Sufficient statistics: Generality

- Example 3: Now consider an ad-valorem tax τ on all products except numeraire.
- Social welfare function is now

$$W(\tau) = \underbrace{\max_x u(x) + (Z + T - p(\tau)\cdot x - \tau p(\tau)\cdot x)}_{\text{Money-metric consumer utility}} + \underbrace{\max_x p(\tau)\cdot x - c(x)}_{\text{Firm profits}}.$$

$$\begin{aligned}\frac{dW(\tau)}{d\tau} &= \frac{d[\tau p(\tau)\cdot x]}{d\tau} - \frac{d[\tau p(\tau)]}{d\tau}\cdot x \\ &= \tau p(\tau)\cdot\frac{dx}{d\tau} \\ &\approx \tau\frac{dE(\tau)}{d\tau},\end{aligned}$$

where $E(\tau)$ is pre-tax expenditure on all goods $1,\ldots,J$.

- Depends only on agg. elasticity, not substitution patterns within taxed group.

Sufficient statistics: Application

- Some notes on applying this sufficient statistics formula.
- 1. dx_1/dt vs. $\partial x_1/\partial t$: which one does reduced-form study measure?
  - Reduced-form studies often measure effect of tax holding other prices fixed, $\approx \partial x_1/\partial t$.
  - E.g., diff-in-diff comparing tax introduced in one county relative to control group $\Rightarrow$ not capturing GE effects of how other prices would change if tax were introduced nationwide.
  - Lesson: Experiment used to identify elasticity must match counterfactual of interest.

Sufficient statistics: Application

- 2. How does one back out dx_1/dt from an observed, non-infinitesimal policy change?
  - For experiment moving tax from t_1 to t_2, observe $\frac{\Delta x_1}{\Delta t}$ instead of $\frac{dx_1(t)}{dt}$.
  - One option: Bound change in welfare. Since:

$$\Delta W = \int_{t_1}^{t_2} t \frac{dx_1(t)}{dt} dt. \qquad \Rightarrow \qquad t_2 \frac{\Delta x_1}{\Delta t} \leq \frac{\Delta W}{\Delta t} \leq t_1 \frac{\Delta x_1}{\Delta t}.$$

  - Second option: Second-order approximation,

$$\Delta W \approx \frac{t_1 + t_2}{2} \Delta x_1.$$

- 3. How does one predict welfare changes outside of sample of observed t?
  - Heckman and Vytacil (2005) recommend researchers estimate schedule of marginal treatment effects and integrate that distribution over desired policy-relevant range.
  - In this setting, use dx_1/dt over observed policy support.
  - Extrapolation for out-of-sample elasticity using functional form?

Sufficient statistics: Application

- 4. How to evaluate the assumptions made to derive sufficient statistics formula?
  - Assumptions implicit in our derivation of Harberger formula.
  - (1) No preexisting distortions, (2) consumers make decisions based only on total prices ($p+t$), (3) consumers and firms maximize, (4) costless implementation and rebates.

Sufficient statistics: Application

- 4. How to evaluate the assumptions made to derive sufficient statistics formula?
  - Assumptions implicit in our derivation of Harberger formula.
  - (1) No preexisting distortions, (2) consumers make decisions based only on total prices ($p+t$), (3) consumers and firms maximize, (4) costless implementation and rebates.

TABLE 3— EFFECT OF POSTING TAX-INCLUSIVE PRICES: *DDD* ANALYSIS OF MEAN QUANTITY SOLD

Period	Control categories	Treated categories	Difference
Panel A. Treatment store			
Baseline (2005:1–2006:6)	26.48	25.17	−1.31
	(0.22)	(0.37)	(0.43)
	[5,510]	[754]	[6,264]
Experiment (2006:8–2006:10)	27.32	23.87	−3.45
	(0.87)	(1.02)	(0.64)
	[285]	[39]	[324]
Difference over time	0.84	−1.30	$DD_{TS} = -2.14$
	(0.75)	(0.92)	(0.68)
	[5,795]	[793]	[6,588]

Figure: Chetty, Looney, and Kroft (2009).

  - Another approach: identify one set of structural params. consistent with elasticities.

Deadweight loss: With tax salience (Chetty, Looney & Kroft, 2009)

- Even with behavioral agents, may be possible to generate sufficient statistics.
- Consider Chetty, Looney, and Kroft (2009) on tax salience.
- Suppose demand for x now depends on both p and t (and assume constant returns). Social welfare:
$$W(t) = u(x) + Z - (p+t)x(p,t) + T(t).$$
- Behavioral agent $\Rightarrow$ Envelope theorem no longer applies. So taking total derivative,
$$\frac{dW}{dt} = \left[u'(x) - p\right]\frac{dx}{dt}.$$
For neoclassical agent, $u'(x) = p + t$ from FOC.

Deadweight loss: With tax salience (Chetty, Looney & Kroft, 2009)

$$\frac{dW}{dt} = \left[u'(x) - p\right] \frac{dx}{dt}.$$

- Suppose we are willing to make one further assumption that when $t = 0$, behavioral agent chooses same x as rational agent.

- Since $u'(x(p,0)) = p$, let $P(x) = x^{-1}(p,0)$. Then we can write:

$$\frac{dW}{dt} = [P(x) - p] \frac{dx}{dt} \approx \frac{(dx/dt)t}{dx/dp} \frac{dx}{dt} = \theta t \frac{dx}{dt},$$

where $\theta = \frac{dx}{dt} / \frac{dx}{dp}$ measures degree of underreaction to tax relative to price.

Deadweight loss: With tax salience (Chetty, Looney & Kroft, 2009)

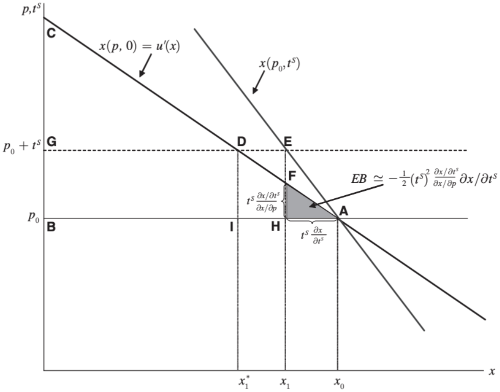


FIGURE 4. EXCESS BURDEN WITH NO INCOME EFFECT FOR GOOD x $(\partial x/\partial Z = 0)$

Notes: This figure illustrates the deadweight cost of introducing a tax t^s levied on consumers when $\partial x/\partial Z = 0$ and producer prices are fixed. The figure plots two demand curves: the price-demand curve $x(p,0)$, which shows how demand varies with the pretax price of the good; and the tax-demand curve $x(p_0, t^S)$, which shows how demand varies with the tax. The figure is drawn assuming $|\partial x/\partial t^S| \leq |\partial x/\partial p|$, consistent with the empirical evidence. The tax reduces demand from x_0 to x_1. The consumer's surplus after the implementation of the tax is given by triangle *DGC* minus triangle *DEF*. The revenue raised from the tax corresponds to the rectangle *GBEH*. The change in total surplus—government revenue plus consumer surplus—equals the shaded triangle *AFH*.

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples
- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples

- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Elasticity of taxable income (Feldstein 1999)

- Income taxes can affect labor supply, but also choice of training, effort, occupation.
- Also can affect choice to avoid or evade taxes.
- For estimating effect of an increase in income tax rate on deadweight loss, do we need to estimate each of these margins individually?

Elasticity of taxable income (Feldstein 1999)

- Income taxes can affect labor supply, but also choice of training, effort, occupation.
- Also can affect choice to avoid or evade taxes.
- For estimating effect of an increase in income tax rate on deadweight loss, do we need to estimate each of these margins individually?
- Feldstein (1999) approach: Suppose individual makes J labor supply choices.
- Total taxable income $=$ income earned from each choice minus sheltered earnings e,

$$\mathrm{TI} = \sum_j w_j x_j - e.$$

- Assume utility linear in consumption for simplicity,

$$u(c, x, e) = c - \underbrace{g(e)}_{\text{Sheltering cost}} - \underbrace{\sum_j \psi_j(x_j)}_{\text{Labor disutility}}, \qquad \text{where} \qquad c = (1-t)\mathrm{TI} + e.$$

Elasticity of taxable income (Feldstein 1999)

- Social welfare function

$$W(t) = \max_{x} \left[(1-t)\mathrm{TI} + e - g(e) - \sum_{j=1}^{J} \psi_j(x_j) \right] + t \cdot \mathrm{TI}.$$

- Taking derivative and using Envelope condition,

$$\frac{dW(t)}{dt} = t\frac{d\mathrm{TI}}{dt}.$$

  Response of taxable income to tax rate is sufficient statistic!

- Moreover, data on taxable income easily available from tax records (especially compared to information about tax evasion / avoidance).

Elasticity of taxable income (Feldstein 1999)

TABLE 2

ESTIMATED ELASTICITIES OF TAXABLE INCOME WITH RESPECT TO NET-OF-TAX RATES

Taxpayer Groups Classified by 1985 Marginal Rate	Net of Tax Rate (1)	Adjusted Taxable Income (2)	Adjusted Taxable Income Plus Gross Loss (3)
	Percentage Changes, 1985–88		
1. Medium (22–38)	12.2	6.2	6.4
2. High (42–45)	25.6	21.0	20.3
3. Highest (49–50)	42.2	71.6	44.8
	Differences of Differences		
4. High minus medium	13.4	14.8	13.9
5. Highest minus high	16.6	50.6	24.5
6. Highest minus medium	30.0	65.4	38.4
	Implied Elasticity Estimates		
7. High minus medium		1.10	1.04
8. Highest minus high		3.05	1.48
9. Highest minus medium		2.14	1.25

NOTE.—The calculations in this table are based on observations for married taxpayers under age 65 who filed joint tax returns for 1985 and 1988 with no age exemption in 1988. Taxpayers who created a subchapter S corporation between 1985 and 1988 are eliminated from the sample.

Figure: Feldstein (1995) analysis of Tax Reform Act of 1986.

- Feldstein (1995) estimates dTI/dt using Tax Reform Act of 1986.
- Large reductions in marginal tax rates.
- Larger increases in taxable income for groups whose tax rates fell more.
- Implied elasticities suggest DWL up to \$2 per \$1 of tax revenue raised on highest-income individuals.
- Subsequent work has tended to find smaller dTI/dt, with active debate.

Elasticity of taxable income (Feldstein 1999)

$$W(t) = \max_{x} \left[(1-t)\mathrm{TI} + e - g(e) - \sum_{j=1}^{J} \psi_j(x_j) \right] + t \cdot \mathrm{TI}.$$

- What does agent maximizing imply about behavior? Recall utility:

$$u(x, e; w, t) = (1-t) \left[\sum_j w_j x_j - e \right] + e - g(e) - \sum_j \psi_j(x_j).$$

- FOC with respect to e is $g'(e) = t$.

- Chetty (2009): What if marginal social cost of tax avoidance not equalized to tax rate?
  - Social vs. private: Some evasion costs are transfers (e.g., fines for tax evasion).
  - Behavioral: individuals overestimate true penalties for evasion.

Elasticity of taxable income (Feldstein 1999)

- Chetty (2009) shows that more general case takes the form,

$$\frac{dW}{dt} = t\left[\frac{g'(e(t))}{t}\frac{d\mathrm{TI}}{dt} + \left(1 - \frac{g'(e(t))}{t}\right)\frac{d\mathrm{LI}}{dt}\right],$$

  where $\mathrm{LI} = \sum_j w_j x_j$ is total earned income.

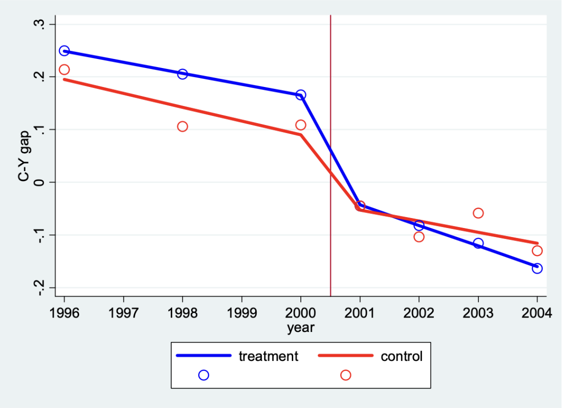
- Feldstein (1999) formula will overstate deadweight losses if $\frac{d\mathrm{LI}}{dt} < \frac{d\mathrm{TI}}{dt}$ and if $g'(e)$ small relative to t.

- How to estimate $g'(e)$?

- Gorodnichenko, Martinez-Vazquez, and Peter (2009) note that if real resource costs of tax evasion are high, that should show up as lower consumption.

Elasticity of taxable income (Feldstein 1999)

Figure 1: Consumption-Income Gap Dynamics for Treatment and Control Groups



Notes: The figure shows annual means of the consumption-income gap for treatment and control groups after controlling for observable characteristics and household fixed effects. Treatment and control group are defined on the basis of post-reform contractual earnings.

Figure: Gorodnichenko, Martinez-Vazquez & Peter (2009).

- Use Russian tax reform in 2001.
- Gap between consumption and income indicates low values of $g'(e)$.
- Moreover, $d\mathrm{TI}/dt$ is substantial relative to $d\mathrm{LI}/dt$.

Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples

- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

Prices vs. quantities (Weitzman 1974)

- Should policymakers should implement a desired intervention with quotas or taxes?
- Suppose quantity q of activity has benefits $B(q)$, produced with cost $C(q)$.
- Assume benefits concave, costs convex, and some crossing point in B' and C'.
- If planner knows each of these functions exactly, simply wants to choose q^* s.t.

$$B'(q^*) = C'(q^*).$$

  Optimal price $p^* = C'(q^*)$.

- With complete certainty, equivalent to announce price p^*, and have producers maximize profits

$$p^* q - C(q).$$

Prices vs. quantities (Weitzman 1974)

- Suppose planner now uncertain about exact functions:

$$B(q,\eta), \qquad \text{and} C(q,\theta),$$

  where η and θ are random variables.

- Ideally, would be able to specify quantity or price in terms of realizations of η and θ,

$$B_1(q^*(\theta,\eta),\eta) = C_1(q^*(\theta,\eta),\theta) = p^*(\theta,\eta).$$

- If planner must commit to price or quantity before realizing signal, which is optimal?

- Suppose planner sets quantity. Then q^{opt} solves

$$q^{\text{opt}} = \arg\max_q \mathbb{E}\left[B(q,\eta) - C(q,\theta)\right].$$

$$\Rightarrow \mathbb{E}\left[B_1\left(q^{\text{opt}},\eta\right)\right] = \mathbb{E}\left[C_1\left(q^{\text{opt}},\theta\right)\right].$$

Prices vs. quantities (Weitzman 1974)

- Suppose planner instead sets price p. Let $\mathcal{Q}(p,\theta)$ denote quantity that results:

$$\mathcal{Q}(p,\theta) = \arg\max_{q} pq - C(q,\theta).$$

$$\Rightarrow \; p = C_1\left(\mathcal{Q}(p,\theta),\theta\right).$$

- Planner thus chooses price to maximize

$$p^{\text{opt}} = \arg\max_{p} \mathbb{E}\left[B\left(\mathcal{Q}(p,\theta),\eta\right) - C\left(\mathcal{Q}(p,\theta),\theta\right)\right].$$

$$\Rightarrow \; p^{\text{opt}} = \frac{\mathbb{E}\left[B_1\left(\mathcal{Q}(p,\theta),\eta\right)\mathcal{Q}_1(p,\theta)\right]}{\mathbb{E}\left[\mathcal{Q}_1(p,\theta)\right]}.$$

- Given realizations of both η and θ, neither q^{opt} or $\mathcal{Q}(p^{\text{opt}},\theta)$ will generally be optimal. So which is better?

Prices vs. quantities (Weitzman 1974)

- Consider a second-order approximation of cost functions around q^{opt} and $\theta = 0$.

$$
\begin{aligned}
C(q,\theta) &\approx C\left(q^{\text{opt}},0\right) + C_1 dq + C_2 d\theta + C_{12} d\theta\, dq + \frac{1}{2} C_{11} dq^2 + \frac{1}{2} C_{22} d\theta^2 \\
&= C\left(q^{\text{opt}},0\right) + C_2 d\theta + \frac{1}{2} C_{22} d\theta^2 + [C_1 + C_{12} d\theta]\left(q - q^{\text{opt}}\right) + \frac{1}{2} C_{11}\left(q - q^{\text{opt}}\right)^2.
\end{aligned}
$$

$$
B(q,\eta) \approx B\left(q^{\text{opt}},0\right) + B_2 d\eta + \frac{1}{2} B_{22} d\eta^2 + [B_1 + B_{12} d\eta]\left(q - q^{\text{opt}}\right) + \frac{1}{2} B_{11}\left(q - q^{\text{opt}}\right)^2.
$$

- This yields first derivatives,

$$
\begin{aligned}
C_1(q,\theta) &= C_1 + C_{12} d\theta + C_{11}\left(q - q^{\text{opt}}\right), \\
B_1(q,\theta) &= B_1 + B_{12} d\eta + B_{11}\left(q - q^{\text{opt}}\right).
\end{aligned}
$$

- Since q^{opt} chosen to equate $\mathbb{E}\left[B_1\left(q^{\text{opt}},\eta\right)\right] = \mathbb{E}\left[C_1\left(q^{\text{opt}},\theta\right)\right]$, we know that $C_1 = B_1$.

Prices vs. quantities (Weitzman 1974)

- We want to know how $\mathcal{Q}(p^{\text{opt}}, \theta)$ will compare to q^{opt}, and whether that is beneficial.

- Recall that $\mathcal{Q}(p^{\text{opt}}, \theta)$ was defined implicitly as

$$p = C_1\left(\mathcal{Q}\left(p,\theta\right),\theta\right) = C_1 + C_{12}d\theta + C_{11}\left(\mathcal{Q}\left(p,\theta\right) - q^{\text{opt}}\right).$$

  Using,

$$p^{\text{opt}} = \frac{\mathbb{E}\left[B_1\left(\mathcal{Q}\left(p,\theta\right),\eta\right)\mathcal{Q}_1\left(p,\theta\right)\right]}{\mathbb{E}\left[\mathcal{Q}_1\left(p,\theta\right)\right]},$$

  we find

$$p^{\text{opt}} = B_1 + \frac{B_{11}}{C_{11}}\left(p^{\text{opt}} - C_1\right) = C_1, \qquad \Rightarrow \qquad \mathcal{Q}\left(p^{\text{opt}},\theta\right) = q^{\text{opt}} - \frac{C_{12}}{C_{11}}d\theta.$$

  If $d\theta > 0$ and $C_{12} > 0$, realized costs high, and production lower given fixed p^{opt}.

Prices vs. quantities (Weitzman 1974)

- What is expected benefit of using prices vs. quantities?

$$\Delta = \mathbb{E}\left[B\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right),\eta\right) - B\left(q^{\text{opt}},\eta\right)\right] - \mathbb{E}\left[C\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right),\theta\right) - C\left(q^{\text{opt}},\theta\right)\right].$$

- Second-order approximations have terms that scale in $(\mathcal{Q} - q^{\text{opt}})$ and $(\mathcal{Q} - q^{\text{opt}})^2$.

Prices vs. quantities (Weitzman 1974)

- What is expected benefit of using prices vs. quantities?

$$\Delta = \mathbb{E}\left[B\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right),\eta\right) - B\left(q^{\text{opt}},\eta\right)\right] - \mathbb{E}\left[C\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right),\theta\right) - C\left(q^{\text{opt}},\theta\right)\right].$$

- Second-order approximations have terms that scale in $(\mathcal{Q} - q^{\text{opt}})$ and $(\mathcal{Q} - q^{\text{opt}})^2$.

- Since $\mathcal{Q}\left(p^{\text{opt}},\theta\right) = q^{\text{opt}} - \frac{C_{12}}{C_{11}}d\theta$, only latter matter in expectation. So:

$$\begin{aligned}
\Delta &= \mathbb{E}\left[\frac{1}{2}B_{11}\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right) - q^{\text{opt}}\right)^2 + \frac{1}{2}C_{11}\left(\mathcal{Q}\left(p^{\text{opt}},\theta\right) - q^{\text{opt}}\right)^2\right] \\
&= \mathbb{E}\left[\frac{1}{2}B_{11}\left(\frac{C_{12}}{C_{11}}d\theta\right)^2 + \frac{1}{2}C_{11}\left(\frac{C_{12}}{C_{11}}d\theta\right)^2\right] \\
&= \frac{C_{12}^2}{2C_{11}^2}\left[B_{11} + C_{11}\right] Var\left(d\theta\right).
\end{aligned}$$

Prices vs. quantities (Weitzman 1974)

$$\Delta = \frac{C_{12}^2}{2C_{11}^2}\left[B_{11} + C_{11}\right] Var\left(d\theta\right).$$

- Sufficient statistic for whether to regulate prices or quantities: Is $B_{11} + C_{11} > 0$?
- I.e., is curvature in benefits (concave) greater or less than curvature in costs.
- Suppose benefits sharply curved. Why would we prefer to set quantities?

Prices vs. quantities (Weitzman 1974)

$$\Delta = \frac{C_{12}^2}{2C_{11}^2}\left[B_{11} + C_{11}\right] Var\left(d\theta\right).$$

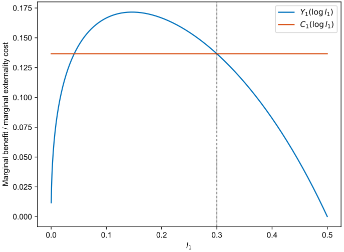
- Sufficient statistic for whether to regulate prices or quantities: Is $B_{11} + C_{11} > 0$?
- I.e., is curvature in benefits (concave) greater or less than curvature in costs.
- Suppose benefits sharply curved. Why would we prefer to set quantities?
- Suppose costs sharply kinked. Why does $\Delta \to 0$?

Prices vs. quantities (Weitzman 1974)

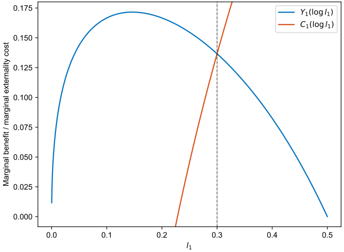
$$\Delta = \frac{C_{12}^2}{2C_{11}^2}\left[B_{11} + C_{11}\right] Var\left(d\theta\right).$$

- Sufficient statistic for whether to regulate prices or quantities: Is $B_{11} + C_{11} > 0$?
- I.e., is curvature in benefits (concave) greater or less than curvature in costs.
- Suppose benefits sharply curved. Why would we prefer to set quantities?
- Suppose costs sharply kinked. Why does $\Delta \to 0$?
- Why does Δ depend on $Var(d\theta)$ and not on $Var(d\eta)$?

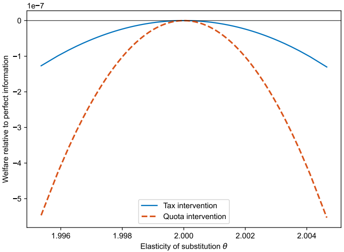
Prices vs. quantities (Weitzman 1974)



(a) Linear costs.



(b) Convex costs.



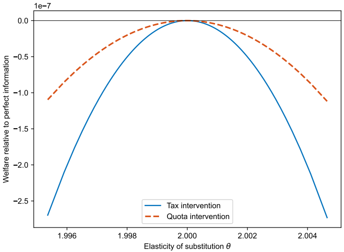


Table of Contents

Presentation tips

Structural, reduced-form, and sufficient statistics approaches

Deadweight loss from a tax

Sufficient statistics: Other examples
- Elasticity of taxable income (Feldstein 1999)
- Prices vs. quantities (Weitzman 1974)

Almost sufficient statistics

“Almost sufficient” statistics

- Mortgage refinancing plays important role in transmission of monetary policy.
- Berger, Milbradt, Tourre, and Vavra (2021) consider whether this channel creates path-dependent effects of monetary policy.
- Idea: Rate cuts encourage borrowers to refinance, but only if they hadn’t already locked in lower rates before.
- Thus, path of previous policy decisions affects current “policy space.”
- Complicated household problem: inattention, fixed costs of refinancing.
- But fraction of outstanding loans with mortgage rates above current market rate (“*frac*”) is roughly sufficient to discipline refinancing response to monetary shock.

“Almost sufficient” statistics: Berger et al. (2021)

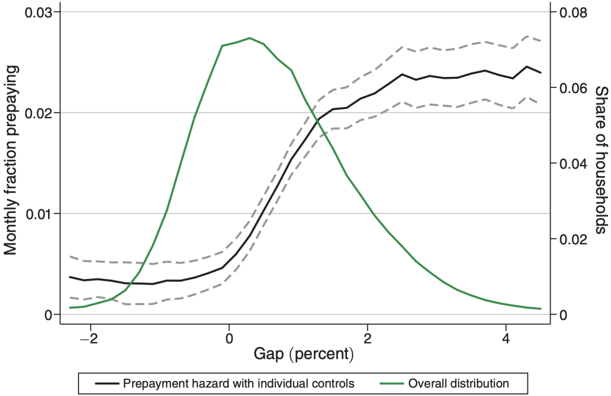


FIGURE 1. PREPAYMENT HAZARD WITH INDIVIDUAL CONTROLS

- Loan originations and mortgage servicing records for 180M loans from Black Knight McDash, 1992–2017.
- Likelihood of repaying in each month.
- Rate gap $=$ interest rate on outstanding loan $-$ predicted rate for new fixed-rate loan given borrower characteristics, loan-to-value.

“Almost sufficient” statistics: Berger et al. (2021)

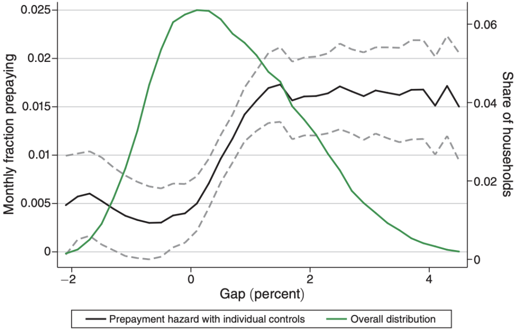


Figure 2. Prepayment Hazard with Individual Controls: Unconstrained Households

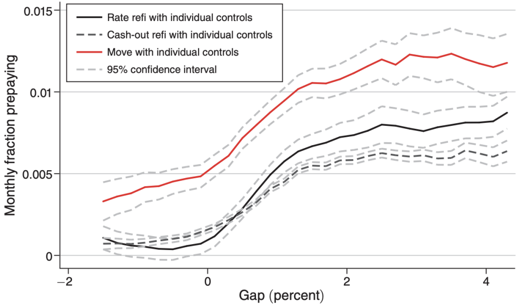


Figure 3. Prepayment Hazard Decomposition with Individual Controls

- Are low refinance rates driven by lender constraints? Doesn’t appear so.
- Rate incentives seem important for all refinancing decisions, even for cash-out refi.

"Almost sufficient" statistics: Berger et al. (2021)

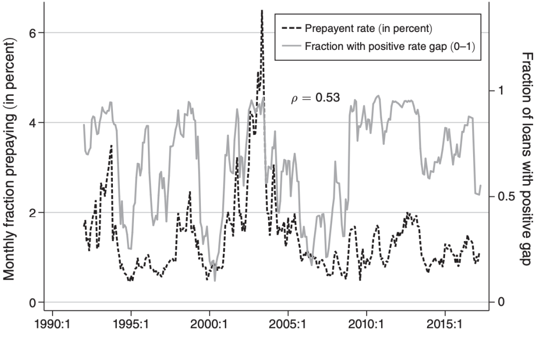


Figure 5. Prepayment versus Fraction with Positive Rate Gap Time Series

Notes: Figure shows the fraction of loans in McDash Performance data with positive rate gaps in each month as well as the fraction of loans prepaying in each month. The time sample is 1992:1–2017:4.

- "The fraction of loans with positive rate gaps in the data, $frac > 0$, has key predictive power for aggregate prepayment rates."
- "If the empirical hazard was an exact step function at 0, then $frac > 0$ would fully summarize all information about how rate incentives affect aggregate prepayment."

“Almost sufficient” statistics: Berger et al. (2021)

TABLE 1—PREDICTIONS OF ALTERNATIVE SUMMARIES FOR RATE INCENTIVES

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
frac $>$ 0	2.051						
	(0.386)						
frac $>$ 50 bps		1.989					
		(0.389)					
frac $>$ 100 bps			2.249				
			(0.450)				
Mean gap				0.759			
				(0.132)			
frac $>$ \$250					1.926		
					(0.384)		
frac $>$ \$500						1.844	
						(0.399)	
frac $>$ \$1,000							1.662
							(0.437)
Constant	0.0416	0.365	0.673	0.993	0.304	0.516	0.941
	(0.196)	(0.145)	(0.107)	(0.0558)	(0.156)	(0.131)	(0.0913)
Adjusted R^2	0.282	0.277	0.246	0.263	0.274	0.247	0.142
Observations	304	304	304	304	304	304	304
Date range	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4	92-17m4

Notes: Newey-West standard errors in parentheses. *frac* $>$ 50 bps is the fraction of loans with gaps greater than 50 bps and *frac* $>$ 100 bps is the fraction with gaps greater than 100 bps. Mean gap is the average gap in a month. *frac* $>$ \$250 (\$500, \$1,000) is the fraction of loans with annual savings greater than \$250 (\$500, \$1,000), which we compute by multiplying the current gap times the outstanding balance. Prepayment fractions are measured in month $t+1$ while rate incentives and LTV are measured in month t, since McDash data measure origination not application and there is a one-to-two-month lag from application to origination.

- Regression of prepayment rate on monthly fraction of loans in full set of bins in Figure 1 yields $R^2 = 0.305$.
- Close to nonparametric description of entire rate gap distribution.
- Thus, *frac* > 0 captures $0.282/0.305 \approx 92.5\%$ of information in full distribution relevant for predicting prepayment.

“Almost sufficient” statistics: Berger et al. (2021)

- Digression: How do we know if mortgage refinancing channel important for aggregate effect of monetary policy?
- Back-of-the-envelope calculation: Overall level of residential mortgage debt $\approx$ \$12.6T.
- How much in savings would refinancing lead to with 100bps reduction in mortgage rates for one year?
  - With 35.4% rate of prepayment / year (2003), $\Rightarrow$ \$44.7B in savings.
  - With 9.2% rate of prepayment / year (2000), $\Rightarrow$ \$11.6B in savings.
  - Potential size of savings varies by 4x given “policy space”!
- With fixed rates, annual payment reductions accrue for life of mortgage.
  - \$44.7B annual reduction in payments has present value of around \$300B.
  - With 128.5M households, this implies present value of \$2,320 / household.
  - If rate reductions spur equity extraction, even more disposal income effect.

“Almost sufficient” statistics: Berger et al. (2021)

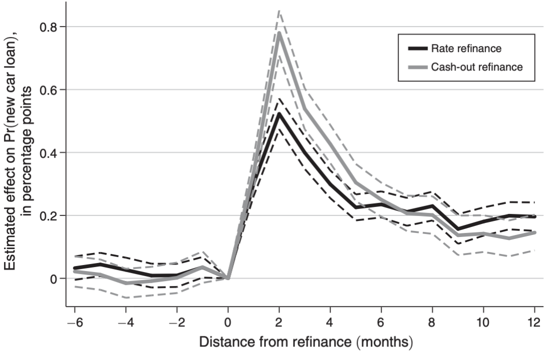


FIGURE 6. INDIVIDUAL LEVEL RESPONSE OF CAR PURCHASE TO REFINANCING

Notes: Figure shows coefficients on month indicators (relative to refinancing month) from regressing a new car loan indicator on household fixed effects, calendar-month fixed effects, and months-from-refinancing indicators, interacted with whether the refinancing involves cash-out. Dashed lines are 95 percent confidence intervals with standard errors two-way clustered by household and calendar month. Our indicator for a new car loan is an increase in auto loan balances of $5,000 or more.

- Direct evidence: Increase P(purchase car) by 3.16pp in 12 months after rate refinance.

“Almost sufficient” statistics: Berger et al. (2021)

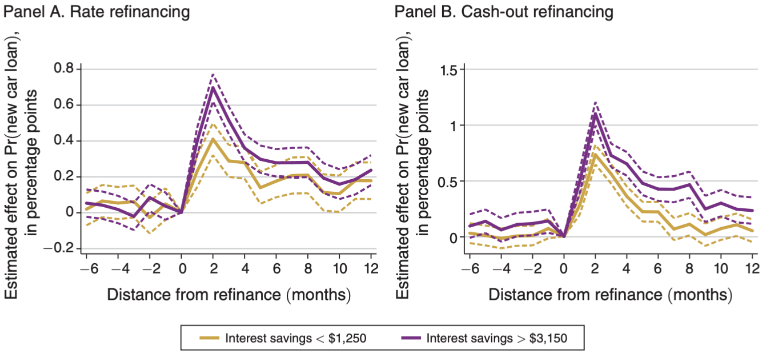


FIGURE 7. HETEROGENEITY BY ANNUAL INTEREST RATE SAVINGS

Note: This figure repeats the analysis in Figure 6 but splitting also by the highest and lowest quartile of interest rate savings.

- Direct evidence: Level of rate savings appear to matter.

“Almost sufficient” statistics: Berger et al. (2021)

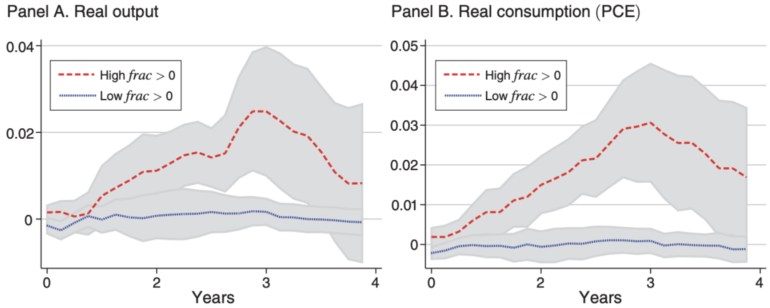


Figure 8. Response to Identified Monetary Policy Shocks

Notes: Figure shows the response of log real output and nondurable consumption to the identified monetary shocks from Romer and Romer (2004), extended through 2007 by Wieland and Yang (2020). Gray areas are 90 percent confidence intervals based on Newey-West standard errors. The red-dashed line indicates the effect of a monetary policy shock when $frac > 0$ is above its median value of 0.596. The blue-dotted line shows the effect when $frac > 0$ is below its median value.

- Cross-region (not shown) and time series evidence on response varying with $frac > 0$.

“Almost sufficient” statistics: Berger et al. (2021)

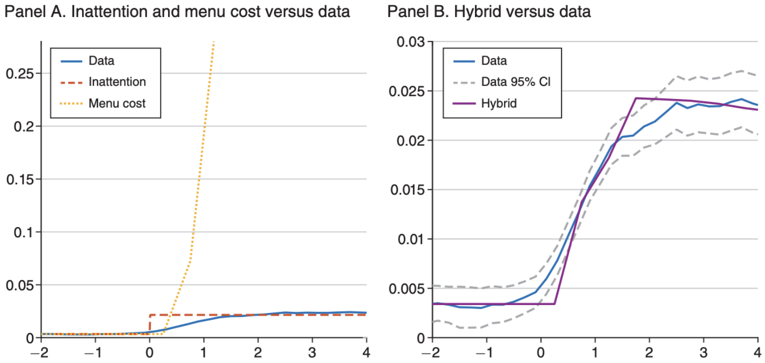


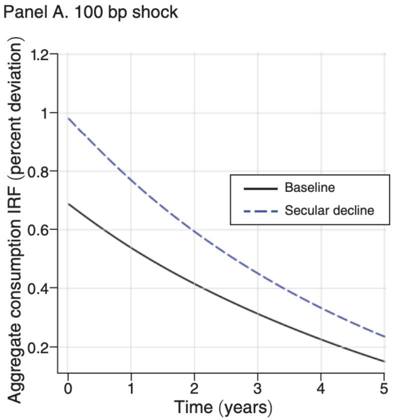
FIGURE 10. PREPAYMENT HAZARDS

- Structural model: Inattention or menu costs alone can’t generate prepayment hazard.
- Hybrid of Calvo (inattention) and menu cost features can rationalize data.

“Almost sufficient” statistics: Berger et al. (2021)

- If adjustment friction is Calvo, then percent of households that adjust is simply $(\textit{frac} > 0) \times \chi_c$, where χ_c is the Calvo hazard.
- If adjustment friction is menu cost, then for infinitesimal shock, effect of changes in the market mortgage rate also depends on $\textit{frac} > 0$!
  - Analogous to result from Caballero and Engel (2007).
  - For infinitesimal shock, state dependence only occurs at gap of zero.
  - Extensive margin of refinancing plays no role in response to small shocks.
- So for theoretical reasons, $\textit{frac} > 0$ also “almost sufficient” statistic.

“Almost sufficient” statistics: Berger et al. (2021)



- Simulations of monetary policy path dependence.
- Secular decline in rates naturally generates higher *frac* > 0 and more “policy space.”
- Substantial differences in consumption IRF depending on previous rates.