Sufficient Statistics for Welfare Analysis

A Bridge Between Structural and Reduced-form Methods

Chetty Raj

Presenter: Qinzhu Sun

December 2, 2021

Overview

Motivation

Two competing paradigms for policy evaluation and welfare analysis:

Structural approach
- **Idea:** specify complete models of economic behavior and estimates or calibrates the primitives; simulate the effects of changes in policies and the economic environment on behavior and welfare.
- **Critique:** hard to identify all primitive parameters in an empirically compelling manner because of endogeneity.

Reduced-form approach
- **Idea:** identify causal effects mainly using research designs that exploit quasi-experimental variation.
- **Critique:** the estimates are not policy-invariant/ deep parameters and therefore have limited relevance for policy and welfare analysis.

Table 1 Recent examples of structural, reduced-form, and sufficient-statistic studies

	Structural	Reduced form	Sufficient statistic
Taxation	Hoynes 1996	Eissa & Liebman 1996	Diamond 1998
	Keane & Moffitt 1998	Blundell et al. 1998	Feldstein 1999
	Blundell et al. 2000	Goolsbee 2000	Saez 2001
	Golosov & Tsyvinksi 2007	Meyer & Rosenbaum 2001	Goulder & Williams 2003
	Weinzierl 2008	Blau & Khan 2007	Chetty 2009
Social insurance	Rust & Phelan 1997	Anderson & Meyer 1997	Gruber 1997
	Golosov & Tsyvinski 2006	Gruber & Wise 1999	Chetty 2006a
	Blundell et al. 2008	Autor & Duggan 2003	Shimer & Werning 2007
	Einav et al. 2008b	Lalive et al. 2006	Chetty 2008
	Lentz 2009	Finkelstein 2007	Einav et al. 2008a
Behavioral models	Angeletos el at. 2001	Genesove & Mayer 2001	Bernheim & Rangel 2008
	İmrohoroğlu et al. 2003	Madrian & Shea 2001	Chetty et al. 2009
	Liebman & Zeckhauser 2004	Shapiro 2005	
	DellaVigna & Paserman 2005	Ashraf et al. 2006	
	Amador et al. 2006	Chetty & Saez 2009	

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

Sufficient Statistics

A middle ground between the two methods: combine
1) the advantages of reduced-form empirics, *transparent and credible identification*, with
2) an important advantage of structural models, *the ability to make precise statements about welfare*.

- Central concept: identify "sufficient statistics" for welfare analysis that can be estimated using reduced-form methods, rather than deep primitives.

- Even though there are multiple combinations of primitives that are consistent with the inputs to the formulas, all such combinations have the same welfare implications.

Sufficient Statistics

Primitives

ω_1
ω_2
.
.
.
ω_N

Sufficient statistics

$\beta_1(t)$
$\beta_2(t)$

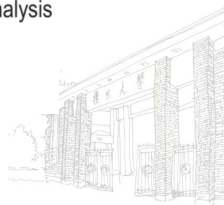**Welfare change**

$\dfrac{dW}{dt}(t)$
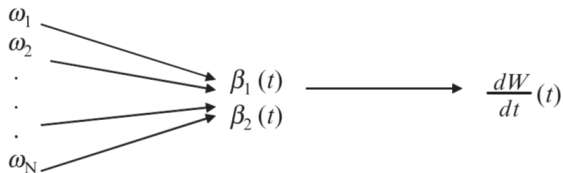
ω = preferences, constraints

ω not uniquely identified

$\beta = f(\omega, t)$
$y = \beta_1 X_1 + \beta_2 X_2 + \varepsilon$

β identified using program evaluation

dW/dt used for policy analysis

Framework

1. A general framework for the derivation of sufficient-statistic formulas for welfare analysis, which shows how envelope conditions from optimization can be used to reduce the set of parameters that need to be identified.

2. A review of several recent papers on tax policy, social insurance, and behavioral public finance.

Marschak's maxim

"...for many decisions (policy problems), only combinations of explicit economic parameters are required – no single economic parameter need be identified." — *Heckman & Vytlacil (2007)*

Pros & Cons of Sufficient-statistic Approach

Pros:

- need to identify fewer parameters (MTE)
- weaker modeling assumptions
- applicable when positive model that generates observed behavior is unknown (e.g. behavioral econ)

Cons:

- new formula must be derived for each question
- easily misapplied because no model evaluation required
- out-of-sample predictions may be less reliable

$\Rightarrow$ Sufficient statistic methods provide useful **complement** to (rather than a substitute for) structural methods.

Setup
Harberger (1964)

Consider a static general equilibrium model:

- An individual is endowed with Z units of numeraire y ;
- Firms convert y into J other consumption goods $x = (x_1, ..., x_J)$;
- $c(x) = \sum_{j=1}^{J} c_j(x_j)$ is total cost of producing x ;
- Perfectly competitive production ;
- The government levies a unit tax t on good 1 ;
- $p = (p_1, ..., p_J)$ is the pretax prices.

Setup
Harberger (1964)

Consumer side:

$$\max_{x,y} u(x_1, ..., x_J) + y \tag{1}$$

$$\text{s.t. } p \cdot x + tx_1 + y = Z$$

Producer side:

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

Market-clearing condition:

$$x^D(p) = x^S(p) \quad \Rightarrow \quad p^* = p(t) \tag{3}$$

Measuring the Efficiency Cost

Harberger (1964)

How to measure the efficiency(deadweight) cost of tax t ?

With quasi-linear utility, the consumer always allocates the lump-sum rebate to consumption of the numeraire y. Social welfare can then be written as the sum of the consumer's utility, producer profits, and tax revenue: [1]

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

(4)

$\Rightarrow$ The efficiency cost is actually dW/dt.

---

[1]Tiny quiz: why does the second "=" hold true?

Measuring the Efficiency Cost

There are two ways to estimate the effect of tax on social welfare $\frac{dW}{dt}$:

First approach

Estimate (or calibrate) a J good demand and supply system to recover the utility function $u(x)$ and cost function $c(x)$. Once u and c are known, directly compute $W(t)$.

Preferences can be recovered using the parametric demand systems; alternatively, one can fit a demand system to the data and then integrate to obtain the expenditure function.

- Challenge: require $2J$ instruments.

Measuring the Efficiency Cost

Harberger (1964)

Second approach ⋆

Differentiating Equation (4) using the envelope theorem:

$$W(t) = \{\max_x u(x) + Z - tx_1 - c(x)\} + tx_1,$$

$$\Rightarrow \frac{dW(t)}{dt} = -x_1 + x_1 + t\frac{dx_1}{dt} = t\frac{dx_1(t)}{dt}. \tag{5}$$

$\Rightarrow \frac{dx_1}{dt}$ is sufficient stat for analyzing the efficiency costs of tax changes.[a]

- Key point: no need to identify the full system of supply and demand curves to calculate ΔW; estimating dx_1/dt directly is enough.
- The loss in social surplus is determined purely by the difference between the agent's willingness to pay for good x_1 and the cost of producing good x_1.

[a]Note that sufficient statistic is *total* derivative, including all GE effects.

Measuring the Efficiency Cost

Harberger (1964)

Harberger Triangle with Fixed Producer Prices
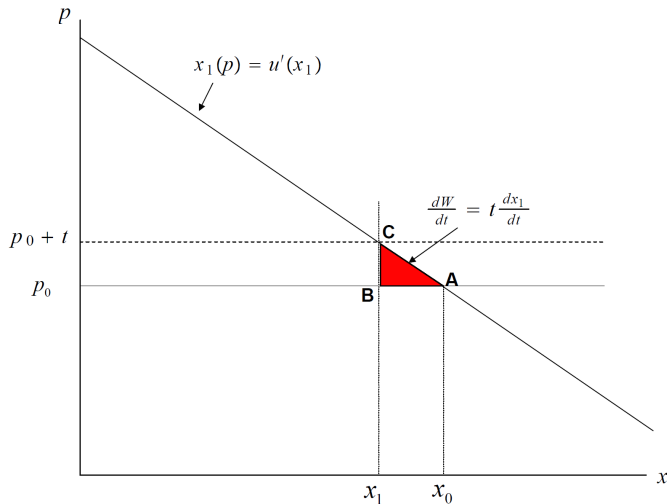
A Brief Summary
Harberger (1964)

Limitations of Harberger's approach:

- It does not permit pre-existing distortions in the other markets; otherwise the spillover effects would have first-order effects on welfare.

- It cannot be used directly to evaluate counterfactual policy changes, such as the imposition of a large new tax on good x_1. [2]

Benefits of Harberger's approach:

- It has greater usage under modern structural models which permits individual heterogeneity and discrete choice. $\Rightarrow$ Two extensive examples since next page...

---

[2] This limitation can be addressed by estimating $dx_1(t)/dt$ for various values of t and making functional-form assumptions to extrapolate out of sample.

Extension 1: Heterogeneity

Individual i is endowed with Z_i units of the numeraire and has utility

$$u^i(x^i) + y. \tag{6}$$

$$\Rightarrow W(t) = \left\{ \sum_{i=1}^{N} \max_{x^i} \left[u^i(x^i) + Z^i - tx_1^i \right] - c(x) \right\} + t \sum_{i=1}^{N} x_1^i \tag{7}$$

$$\Rightarrow \frac{dW(t)}{dt} = -\sum_{i=1}^{N} x_1^i + \sum_{i=1}^{N} x_1^i + t \frac{d \sum_{i=1}^{N} x_1^i}{dt} = t \frac{dx_1(t)}{dt}. \tag{8}$$

So $\frac{dx_1}{dt}$ is a sufficient statistic for the marginal excess tax burden.

- There is no need to characterize the underlying heterogeneity in the population to implement Equation (8);
- Even though each individual diverges in demand elasticity, what matters for government revenue and aggregate welfare is the total change in behavior induced by the tax.
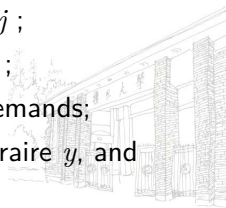
Extension 2: Discrete Choice

Setup:

- Individuals choose one of the J products $\{1, ..., J\}$;
- Each product is characterized by a vector of K observable attributes $x_j = (x_{1j}, ..., x_{Kj})$ and an unobservable attribute ζ_j ;

If agent i chooses product j, his utility is

$$
\begin{aligned}
u_{ij} &= v_{ij} + \varepsilon_{ij}, \\
\text{with } v_{ij} &= Z^i - p_j + \zeta_j + \phi^i(x_j).
\end{aligned}
\tag{9}
$$

Denotation:

- P_{ij} : the probability that individual i chooses option j ;
- $P_j = \sum_i P_{ij}$: total (expected) demand for product j ;
- $P = (P_1, ..., P_J)$: the vector of aggregate product demands;
- Product j is produced using $c_j(P_j)$ units of the numeraire y, and $c(P) = \sum_j c_j(P_j)$.

Extension 2: Discrete Choice

Binary Choice Case

Suppose individuals make binary decisions about whether to buy x_1.

- If he doesn't buy x_1 and spends wealth on y, $u_i = Z^i$;
- If he buys x_1, $u_i = Z^i - p_1 + \zeta_1 + \phi^i(x_1) + \varepsilon_{i1}$.

Denotation:

- $V^i = \zeta_i + \phi^i(x_1) + \varepsilon_{i1}$: individual i's gross valuation of x_1 ;
- $F(V^i)$: the smooth distribution of valuations in the economy;
- EZ : the average level of wealth in the economy;
- $\bar{V}$: the cutoff of V^i to buy x_1.

$$W(t) = \left\{ EZ + \max_{\bar{V}} \int_{\bar{V}}^{\infty} \left[V^i - (p_1 + t) \right] dF(V^i) \right\} + t \int_{\bar{V}}^{\infty} dF(V^i) \quad (10)$$

$$\frac{dW}{dt} = -\left[1 - F(\bar{V}) \right] + \left[1 - F(\bar{V}) \right] + t \frac{d \int_{\bar{V}}^{\infty} dF(V^i)}{dt} = t \frac{dx_1}{dt} \quad (11)$$

Extension 2: Discrete Choice

Multinomial Choice Case

Assume that ε_{ij} has a type 1 extreme value distribution. $\Rightarrow$ The probability that a utility-maximizing individual i chooses product j is
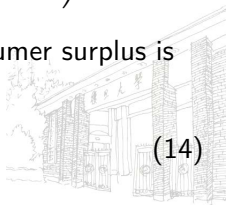
$$P_{ij} = \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \tag{12}$$

and that agent i's expected utility from price vector $p = (p_1, ..., p_J)$ is

$$S_i(p_1, ..., p_J) = E \max(u_{i1}, ..., u_{iJ}) = \log \left(\sum_j \exp(v_{ij}) \right). \tag{13}$$

Aggregating over $i = 1, ..., N$ consumers, (expected) consumer surplus is

$$S = \sum_i \log \left(\sum_j \exp(v_{ij}) \right). \tag{14}$$

Extension 2: Discrete Choice

Multinomial Choice Case

Add producer profits to obtain social welfare:

$$W = \sum_i \log \left(\sum_j \exp(v_{ij}) \right) + p \cdot P - c(P) \tag{15}$$

Sufficient-statistic approaches offer a means of policy analysis that does not require identifying ϕ_i and ζ_j. Suppose the government levies a tax t on good 1, raising its price to $p_1 + t$. The proceeds are returned to agents through a lump-sum transfer T so that y_i becomes $y_i + T$.

$$\Rightarrow \frac{dW(t)}{dt} = \sum_i \left[-\frac{\exp(v_{i1})}{\sum_j \exp(v_{ij})} - \sum_j \frac{dp_j}{dt} \frac{\exp(v_{ij})}{\sum_j \exp(v_{ij})} \right] \\ + \sum_j \frac{dp_j}{dt} P_j + P_1 + t \frac{dP_1}{dt} = t \frac{dP_1(t)}{dt}. \tag{16}$$

Extension 2: Discrete Choice
Multinomial Choice Case

Now suppose that an ad-valorem tax t is levied on all the products except the numeraire good, raising the price of product j to $(1 + \tau)p_j$. Similarly,

$$\frac{dW(\tau)}{d\tau} = \tau \sum_j p_j \frac{dP_j(\tau)}{d\tau} = \tau \frac{dE_P(\tau)}{d\tau}, \tag{17}$$

where $E_p = \sum_j p_j P_j$ denotes total pretax expenditure for the taxed good.

- Finding: The efficiency cost of a tax on all products depends on the aggregate expenditure elasticity for the taxed market; it does not require estimation of the substitution patterns within that market.

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

A Six(Five) Step Rubric

1. Specify model structure (prefs, technology) and social welfare function:

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

2. Write dW/dt in terms of marginal utilities using envelope conditions.

3. Recover marginal utilties from observed choices using comparative statics of model to obtain sufficient stat formula

$$\frac{dW}{dt}(t) = f\left(t, \frac{dx_1}{dt}, \frac{dx_1}{dZ}, \frac{dx_2}{dt}, \frac{dx_2}{dZ}, ...\right).$$

4. Empirical implementation: mapping non-marginal LATE estimates to sufficient statistics

5. Model evaluation: do structural assumptions fit the data?
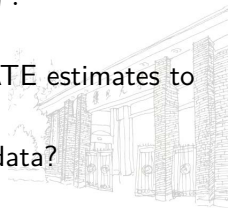
Specify the Model Structure

Consumer decision:

$$\max U(x) \text{ s.t. } G_1(x, t, T) = 0, ..., G_M(x, t, T) = 0. \tag{18}$$

- A unit tax t is levied on choice x_1 and the transfer $T(t)$ is paid in units of x_J ;
- $\{G_1(x, t, T), ..., G_M(x, t, T)\}$ denote the $M < J$ constraints.

$$\Rightarrow W(t) = \max_x U(x) + \sum_{m=1}^{M} \lambda_m G_m(x, t, T). \tag{19}$$

An example of single-agent Harberger model ▸Back

$$U(x) = u(x_1, ..., x_{J-1}) + x_J$$
$$G_1(x, t, T) = T + Z - t x_1 - c(x_1, ..., x_{J-1}) - x_J$$

Express dW/dt in Terms of Multipliers

Using the envelope conditions,

$$\frac{dW}{dt} = \sum_{m=1}^{M} \lambda_m \left\{ \frac{\partial G_m}{\partial T} \frac{dT}{dt} + \frac{\partial G_m}{\partial t} \right\}. \tag{20}$$

- $\frac{dT}{dt}$ is known through the govt's budget constraint;
- $\frac{\partial G_m}{\partial T}$ and $\frac{\partial G_m}{\partial t}$ can be calculated mechanically.

$\Rightarrow$ The critical unknowns are the λ_m multipliers.

Substitute Multipliers by Marginal Utilities

The λ_m multipliers are recovered by exploiting restrictions from the agent's first-order conditions:

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

To simplify the inverse mapping from λ to $u'(\cdot)$, impose as followed:

Assumption 1 (Interchangability Condition)

x_1 and t (x_J and T) enter every constraint interchangeably: [a]

$$\frac{\partial G_m}{\partial t} = k_t(x, t, T) \frac{\partial G_m}{\partial x_1} \qquad \forall m = 1, ..., M,$$

$$\frac{\partial G_m}{\partial T} = -k_T(x, t, T) \frac{\partial G_m}{\partial x_J} \qquad \forall m = 1, ..., M$$

[a] In models with only one constraint per agent, this is satisfied by definition.

Substitute Multipliers by Marginal Utilities

Substitute Assumption 1 into equation (20) to obtain

$$
\begin{aligned}
\frac{dW}{dt} &= \sum_{m=1}^{M} \lambda_m \left\{ -k_T \frac{\partial G_m}{\partial x_J} \frac{dT}{dt} + k_t \frac{\partial G_m}{\partial x_1} \right\} \\
&= -k_T \frac{dT}{dt} \sum_{m=1}^{M} \lambda_m \frac{\partial G_m}{\partial x_J} + k_t \sum_{m=1}^{M} \lambda_m \frac{\partial G_m}{\partial x_1} \qquad (21) \\
&= k_T \frac{dT}{dt} u'(x_J(t)) - k_t u'(x_1(t)).
\end{aligned}
$$

Intuition

- Increasing the tax t is equivalent to reducing consumption of x_1 by k_t units, which reduces the agent's utility by $k_t u'(x_1(t))$.
- The additional transfer that the agent gets from the tax increase is $\frac{dT}{dt} k_T$ units of good x_J, which raises his utility by $k_T \frac{dT}{dt} u'(x_J(t))$.

k_T, k_t, and $\frac{dT}{dt}$ are known based on model specification. $u'(\cdot)$??

Recover Marginal Utilities from Observed Choices

The final step in obtaining an empirically implementable expression for $\frac{dW}{dt}$ is to back out the two marginal utilities.

- Sufficient-statistic studies recover the marginal utilities from choice data using the model structure specified in step 1. There is no canned procedure for this step.
- The usual trick is that the marginal utilities are elements in first-order conditions for various choices. As a result, they can be backed out from the comparative statics of behavior. ▸ Feldstein (1999)

Example: Single-agent Harberger Model ▸ Example

The assumption of no income effects implies $u'(x_J) = 1$. The first-order condition for x_1 leads to $u'(x_1) = p_1 + t$.

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

Empirical Implementation

Suppose the sufficient-statistic formula has the following form:

$$\frac{dW}{dt}(t) = f\left(\frac{\partial x_1}{\partial t}, \frac{\partial x_1}{\partial Z}, t\right). \tag{22}$$

Note 1

The relevant derivatives may require holding different variables fixed depending upon the application.

- E.g. the Harberger formula in Equation (5) measures $\frac{dx_1}{dt}$, which incorporates general equilibrium effects & price changes in all markets.
- In contrast, many reduced-form empirical studies explicitly identifies $\frac{\partial x_1}{\partial t}$, holding prices in other markets constant. Such studies do not recover the sufficient stat of interest for Harberger policy questions.

$\Rightarrow$ The general lesson is that the experiment used to identify the relevant elasticities must be matched to the policy question being asked.

Empirical Implementation

Note 2

- The ideal way to implement Equation (22) to assess the efficiency costs of a discrete policy change is to estimate the inputs as nonparametric functions of the policy instrument t.

- With estimates of $\frac{\partial x_1}{\partial t}(t)$ and $\frac{\partial x_1}{\partial Z}(t)$, one can integrate Equation (22) between any two tax rates t_1 and t_2 to evaluate ΔW. [a]

---
[a]Similar in spirit to Heckman and Vytlacil (2001, 2005): from MTEs to policy-relevant treatment effects.

In most applications, there is insufficient power to nonparametrically estimate $x_1(t)$. Instead, typical reduced-form studies estimate LATE, such as $\frac{\Delta x_1}{\Delta t} = \frac{x_1(t_2) - x_1(t_1)}{t_2 - t_1}$.

Empirical Implementation

To see this, consider the Harberger model $\frac{dW}{dt}(t) = t\frac{dx_1}{dt}(t)$. A researcher who has estimated $\frac{\Delta x_1}{\Delta t}$ has two options.

Option 1: bound the average welfare gain over the observed range.

$$W(t_2) - W(t_1) = \int_{t_1}^{t_2} \frac{dW}{dt} dt = \int_{t_1}^{t_2} t\frac{dx_1}{dt}(t) dt$$
$$\Rightarrow t_1 \frac{\Delta x_1}{\Delta t} > \overline{dW/dt} > t_2 \frac{\Delta x_1}{\Delta t} \tag{23}$$

▸ Harberger Triangle

Option 2: approximate to $x_1(t)$ to calculate $\overline{dW/dt}$. E.g. [3]

$$\frac{dW(t)}{dt} = t\frac{dx_1(t)}{dt} \quad \Rightarrow \quad \overline{dW/dt} \simeq \frac{t_1 + t_2}{2} \frac{\Delta x_1}{\Delta t} \tag{24}$$
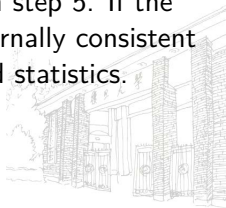
▸ Feldstein (1999)

---

[3]Approximate that dx_1/dt is constant over the observed range.

Model Evaluation

Although sufficient-statistic formulas do not require full specification of the model, they do require some modeling assumptions. It is important to assess the validity of these assumptions to ensure the accuracy of results.

The model can be evaluated in two ways.

1. One can test qualitative predictions that would falsify the assumptions that are central for deriving the sufficient-statistic formula.

2. One should identify at least one vector of structural parameters ω that is consistent with the sufficient statistics estimated in step 5. If the empirical estimates of the sufficient statistics are internally consistent with the model, at least one ω must fit the estimated statistics.

Income Taxation

Feldstein (1999)

Setup:

- An individual makes J labor-supply choices $(x_1, ..., x_J)$ that generate income;
- w_j denotes the wage paid for choice j; $\psi_j(x_j)$ denotes the disutility of labor supply through margin x_j ;
- The agent shelters e of earnings from the tax authority by paying $g(e)$;
- Total taxable income is $TI = \sum_{j=1}^{J} w_j x_j - e$;
- Consumption is then $c = (1 - t) TI + e$.

We can write the model formally as

$$u(c, x, e) = c - g(e) - \sum_{j=1}^{J} \psi_j(x_j)$$

$$T(t) = t \cdot TI$$

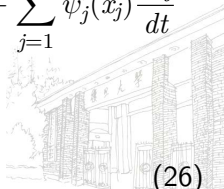$$G_1(c, x, t) = T + (1 - t) TI + e - c.$$

Income Taxation

Feldstein (1999)

Social welfare is [4]

$$W(t) = u(c, x, e) + G_1(c, x, t)$$

$$= \left\{ (1 - t)\, TI + e - g(e) - \sum_{j=1}^{J} \psi_j(x_j) \right\} + t \cdot TI. \tag{25}$$

To calculate the marginal excess burden $\frac{dW}{dt}$, differentiate equation (25):

$$\frac{dW}{dt} = TI + t\frac{dTI}{dt} - TI + (1 - t)\frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt}$$

$$= \frac{dTI}{dt} + \frac{de}{dt}(1 - g'(e)) - \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt}. \tag{26}$$

---

[4] Tiny quiz: where is the Lagrangian multiplier λ?

Income Taxation
Feldstein (1999)

To recover the marginal utilities (▸ step 4), Feldstein exploited the F.O.C.

$$g'(e) = t$$

$$\psi_j'(x_j) = (1-t)w_j \Rightarrow \sum_{j=1}^{J} \psi_j'(x_j)\frac{dx_j}{dt} = \sum_{j=1}^{J}(1-t)w_j\frac{dx_j}{dt} = (1-t)\frac{d(TI+e)}{dt}$$
(27)

Plug equation (27) into equation (26), we obtain [5]

$$\frac{dW(t)}{dt} = t\frac{dTI(t)}{dt}.$$
(28)

$\Rightarrow$ We simply need to measure how taxable income responds to changes in the tax rate to calculate the deadweight cost of income taxation.

---

[5] The result is actually obvious using the envelope conditions.

Income Taxation
Feldstein (1999)

- The main advantage of identifying $dTI(t)/dt$ as a sufficient statistic is that it permits inference about efficiency costs without requiring the identification of the potentially complex effects of taxes on numerous labor supply, evasion, and avoidance behaviors.

- Moreover, data on taxable income are available on tax records, facilitating the estimation of the key parameter dTI/dt.

Example: Feldstein (1995)

Feldstein (1995) implemented Equation (28) by estimating the changes in reported taxable income around the Tax Reform Act of 1986, implicitly using the linear approximation described in ▸ step 5 .

He concluded that the excess burden of taxing high-income individuals is very large, possibly as large as \$2 per \$1 of revenue raised.

Income Taxation
Feldstein (1999)

Model evaluation: Chetty (2009) argued that the marginal social cost of tax avoidance may not be equal to the tax rate at the optimum, violating the F.O.C. that is critical to derive Equation (28).

1. Some of the costs of evasion and avoidance constitute transfers rather than resource costs;
2. There is considerable evidence that individuals overestimate the true penalties for evasion.

Chetty relaxed the $g'(e) = t$ restriction and obtained the following generalization of Feldstein's formula:

$$\frac{dW(t)}{dt} = t\left\{\mu(t)\frac{dTI(t)}{dt} + (1 - \mu(t))\frac{dLI(t)}{dt}\right\}, \quad (29)$$

where $LI = \sum_{j=1}^{J} w_j x_j$ represents total earned income and $\mu(t) = \frac{g'(e(t))}{t}$ measures the gap between social marginal costs of avoidance and tax rate.

Income Taxation
Feldstein (1999)

$$\frac{dW(t)}{dt} = t\left\{\mu(t)\frac{dTI(t)}{dt} + (1 - \mu(t))\frac{dLI(t)}{dt}\right\},$$

- Intuitively, deadweight loss is a weighted average of
    - the taxable income elasticity dTI/dt, and
    - the total earned income elasticity dLI/dt.
- The weight is determined by the resource cost of sheltering.
- If avoidance does not have a large resource cost, changes in e have little efficiency cost, and thus it is only dLI/dt —the real labor-supply response—that matters for deadweight loss.
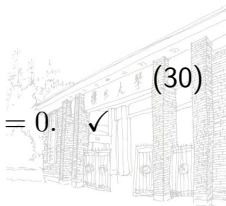
Income Taxation
Saez (2001)

Setup:

- Individuals choose hours of work, l, and have heterogeneous wage rates w distributed according to a distribution $F(w)$;
- Pretax earnings is denoted $z = wl$;
- The govt levies a linear tax τ on earnings above a threshold $\bar{z}$;
- $c(w, \tau)$, $l(w, \tau)$, $z(w, \tau) = wl(w, \tau)$ are the agent's optimal choices.

For a given $\bar{z}$, individuals maximize utility

$$u(c, l) = c - \psi(l)$$

$$\text{s.t. } G_1(c, l) = (1 - \tau) \max(wl - \bar{z}, 0) + \bar{z} - c = 0. \quad \times \tag{30}$$

$$\text{s.t. } G_1(c, l) = (1 - \tau) \max(wl - \bar{z}, 0) + \min(wl, \bar{z}) - c = 0. \quad \checkmark$$

Income Taxation
Saez (2001)

Let

- $z_m(\bar{z}) = E[wl(w,\tau)|z(w,\tau) > \bar{z}]$ denote the mean level of earnings for individuals in the top bracket;
- $\bar{w}$ represent the wage threshold that corresponds to an earnings threshold of $\bar{z}$ when the tax rate is τ: $\bar{w}l(\bar{w},\tau) = \bar{z}$.

$\Rightarrow$ The tax revenue generated by the top bracket tax is $R = \tau(z_m(\bar{z}) - \bar{z})$.

Social planner's objective: maximize a weighted average of individual's utilities, where the weights $\tilde{G}(u)$ are social-welfare weights[6] that reflect the redistributive preferences of the planner.

$$W = \left\{ \int_0^\infty \tilde{G}(u(c(w,\tau), wl(w,\tau))) dF(w) \right\} + \tau(z_m(\bar{z}) - \bar{z}) \quad (31)$$

---

[6]Could be utilitarian or Rawlsian or ...

Income Taxation
Saez (2001)

$$\Rightarrow \frac{dW}{d\tau}(\tau) = -\int_{\bar{w}}^{\infty} \tilde{G}_u(u)(z(w,\tau) - \bar{z})dF(w) + \left[(z_m - \bar{z}) + \tau\frac{dz_m}{d\tau}\right]$$
$$= -(z_m(\bar{z}) - \bar{z})\bar{g} + \left[(z_m(\bar{z}) - \bar{z}) + \tau\frac{dz_m}{d\tau}\right],$$

(32)

where

- $\bar{g} = \int_{\bar{w}}^{\infty} \tilde{G}_u(u)(z - \bar{z})dF(w) / \int_{\bar{w}}^{\infty}(z - \bar{z})dF(w)$ denotes the mean marginal social-welfare weight placed on top-income individuals;
- $\bar{g}$ measures the social value of giving \$1 more income to individuals in the top bracket relative to the value of public expenditure.

$\Rightarrow \frac{dz_m}{d\tau}$, $z_m(\bar{z})$, $\bar{g}$ are together sufficient statistics for the welfare gain of increasing top income-tax rates. & No need to identify preferences ψ or the shape of the skill distribution $F(w)$.

Income Taxation

The disadvantage is that $\frac{dz_m}{d\tau}$, $z_m(\bar{z})$, $\bar{g}$ are endogenous to τ.

$\Rightarrow$ Saez observed that the upper tail of the earnings distribution is well described by a Pareto distribution in US.[7] Hence, equation (32) can be expressed as

$$\frac{dW}{dt} = \frac{1 - \bar{g}}{a - 1}\bar{z} + \tau\frac{dz_m}{d\tau}. \tag{33}$$

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

$$\frac{\tau^*}{1 - \tau^*} = \frac{1 - \bar{g}}{a\varepsilon}, \tag{34}$$

where $\varepsilon = \frac{dz_m}{d(1-\tau)}\frac{1-\tau}{z_m}$ is the taxable income elasticity in the top bracket.

$\Rightarrow$ Equation (34) is an explicit formula for the optimal asymptotic top income-tax rate.

---

[7]A Pareto distribution with parameter a has $\frac{z_m(\bar{z})}{\bar{z}} = \frac{a}{a-1}$ for all $\bar{z}$.
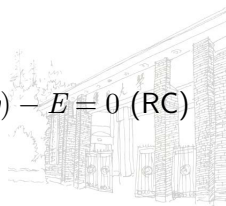
Income Taxation

Saez characterized the optimal tax rate in a nonlinear tax system.

Denotation:

- $\varepsilon(z) = \frac{dz}{d(1-\tau)} \frac{1-\tau}{z}$: the earnings elasticity at income level z ;
- $h(z)$: the density of the earnings distribution at z ;
- $\tilde{G}(u(z))$: the weight that the planner places on an individual;
- $g(Z) = \tilde{G}_u \cdot u_c(z)$ is the marginal social-welfare weight.

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

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

$$\text{s.t. } G_1(c, z, T) = \int_0^\infty z(w, T) \, dF(w) - \int_0^\infty c(w, T) \, dF(w) - E = 0 \text{ (RC)}$$

$$G_2(c, z, T) = (1 - T'(z))w - \psi'(l(w)) = 0 \text{ (IC)}$$

Income Taxation
Saez (2001)

Exploiting envelope conditions and perturbation arguments as above, the optimal tax schedule satisfies the following condition at all z:

$$\frac{T(z)}{1 - T(z)} = \frac{1}{\varepsilon(z) z h(z)} \int_z^\infty (1 - g(z')) h(z') \, dz', \tag{35}$$

which depends on the same 3 parameters as equation (32): the taxable income elasticity $\varepsilon(z)$, the shape of earnings distribution $h(z)$, and the social-welfare weights $g(z)$.

To go further and calculate the optimal tax schedule, more assumptions are needed about the model's structure. (Examples not presented here)

Social Insurance
Baily (1978) and Chetty (2006a)

Setup:

- w_h and w_l are the individual's income in high / low state ($w_l < w_h$);
- A denote wealth;
- c_h and c_l are the individual's consumption in high / low state;
- The probability of being in the high state is $p(e) = e$. The agent can control it by exerting effort e at a cost $\psi(e)$.
- The agent can transfer b_p between states at a cost $q(b_p)$, so that increasing consumption by b_p in the low state requires payment of a premium $\frac{1-e}{e}b_p + q(b_p)$ in the high state.[8]
- The government pays a benefit b in the low state that is financed by an actuarially fair tax $t(b) = \frac{1-e}{e}b$ in the high state.

---

[8]The loading factor $q(b_p)$ can be interpreted as the degree of incompleteness in private insurance.

Social Insurance
Baily (1978) and Chetty (2006a)

$$U(c_l, c_h, e) = eu(c_h) + (1 - e)u(c_l) - \psi(e), \quad t(b) = \frac{1 - e}{e}b$$

$$s.t. \ G_1(c_l, c_h, t) = c_h + \frac{1 - e}{e}b_p + q(b_p) + t - w_h - A = 0 \quad (36)$$

$$G_2(c_l, c_h, t) = c_l - b_p - b - w_l - A = 0$$

$$\Rightarrow W(b) = eu\left(A + w_h - \frac{1 - e}{e}b_p - q(b_p) - t(b)\right)$$

$$+ (1 - e)u(A + w_l + b_p + b) - \psi(e)$$

$$\Rightarrow \frac{dW(b)}{db}$$

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

where $\varepsilon_{1-e,b} = \frac{d(1-e)}{db}\frac{b}{1-e}$ is the elasticity of the probability of being in the bad state $1 - e$ w.r.t. the benefit level b.

Social Insurance
Baily (1978) and Chetty (2006a)

Normalize the welfare gain from a 1 (balanced budget) increase in the size of the government insurance program by the welfare gain from raising the wage bill in the high state by 1:

$$M_W(b) = \frac{\frac{dW}{db}(b)/(1-e)}{\frac{dW}{dw_h}(b)/e} = \frac{u'(c_l) - u'(c_h)}{u'(c_h)} - \frac{\varepsilon_{1-e,b}}{e} \qquad (37)$$

- $\frac{u'(c_l) - u'(c_h)}{u'(c_h)}$ measures the gap in marginal utilities between states, which quantifies the welfare gain from transferring an additional dollar from the high to low state;
- $\frac{\varepsilon_{1-e,b}}{e}$ measures the cost of transferring this additional dollar due to behavioral responses.

$\Rightarrow$ The parameters in equation (37) are sufficient statistics.

Social Insurance
Baily (1978) and Chetty (2006a)

Equation (37) cannot be directly implemented because the gap in marginal utilities remains to be recovered from choice data.

The recent literature has proposed the use of three types of choice data to recover the marginal utility gap:

- consumption (Gruber 1997),
- liquidity and substitution effects in effort (Chetty 2008),
- reservation wages (Shimer and Werning 2007).

Social Insurance
Gruber (1997)

Taking a quadratic approximation to the utility function, Gruber observed that

$$\frac{u'(c_l) - u'(c_h)}{u'(c_l)} = \gamma \frac{\Delta c}{c_h}(b), \tag{38}$$

where $\gamma = -\frac{u''(c_h)}{u'(c_h)} c_h$ is the relative risk aversion and $\Delta c = c_h - c_l$.

He posited that the effect of UI benefits on consumption is linear:

$$\frac{\Delta c}{c_h}(b) = \alpha + \beta b \tag{39}$$

$\Rightarrow$ Combining with equation (37) & (38), we obtain the marginal welfare gain from increasing the benefit level

$$M_W(b) = (\alpha + \beta b)\gamma - \frac{\varepsilon_{1-e,b}}{e}. \tag{40}$$

Social Insurance

Chetty (2008)

Observe that the first-order condition for effort is ▸ Recall

$$\psi'(e) = u(c_h) - u(c_l). \tag{41}$$

Now consider the effect of an exogenous cash grant on effort (next page):

$$\partial e/\partial A = \left\{ u'(c_h) - u'(c_l) \right\} /\psi''(e) \le 0 \tag{42}$$

The effect of increasing the benefit level on effort is

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

$$\Rightarrow M_W(b) = \frac{u'(c_l) - u'(c_h)}{u'(c_h)} - \frac{\varepsilon_{1-e,b}}{e} = \frac{-\partial e/\partial A}{\partial e/\partial A - \partial e/\partial b} - \frac{\varepsilon_{1-e,b}}{e} \tag{43}$$

Liquidity effect de/dA measures completeness of private insurance; moral hazard effect de/dw_h measures efficiency cost of insurance.

Social Insurance

Chetty (2008)

Reduced-form estimates of de/dA

Card, Chetty, and Weber (2007)

Effect of Severance Pay on Unemployment Durations in Austria

Social Insurance
Shimer and Werning (2007)

The probability of finding a job, e, depends on one's decision to accept or reject a wage offer. Offers are drawn from a distribution $F(w)$. If he rejects, he receives $w_l + b$. The agent has no private insurance ($q = \infty$).

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

$$W(b) = eE\left[u(w - t)|w - t > w_l + b\right] + (1 - e)u(w_l + b). \qquad (44)$$
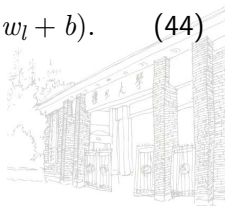
Social Insurance
Shimer and Werning (2007)

Define the agent's reservation wage prior to job search as the wage $\bar{w}_0$.[9]
The reservation wage $\bar{w}_0$ satisfies

$$u(\bar{w}_0 - t) = W(b). \tag{45}$$

The government's problem is to

$$\max W(b) = \max u(\bar{w}_0 - t) \\ \Rightarrow \max \bar{w}_0 - t \tag{46}$$

Differentiation gives a sufficient-statistic formula

$$M_W(b) = \frac{d\bar{w}_0}{db} - \frac{dt}{db} = \frac{d\bar{w}_0}{db} - \frac{1-e}{e}\left(1 + \frac{1}{e}\varepsilon_{1-e,b}\right). \tag{47}$$

---

[9] $\bar{w}_0$ would make the agent indifferent about accepting a job immediately to avoid having to take a random draw from the wage-offer distribution.

Social Insurance
Inefficiencies in Private Insurance

- An important assumption made in all three formulas above is that the choices within the private sector are constrained Pareto efficient.
- In practice, private-insurance contracts are likely to be second-best inefficient as well because of adverse selection and moral hazard in private markets. In this case, the envelope condition invoked is violated because of externalities on the private insurer's budget constraint that are not taken into account by the individual.
- Recent work...

Behavioral Models

There is now considerable reduced-form evidence that individuals' behavior deviates systematically from the predictions of neoclassical perfect optimization models.

- The budding literature on this topic has proposed some structural approaches, primarily in the context of time discounting.
- Another set of studies has modeled the behavioral patterns identified in earlier work[10] and simulated optimal tax policy in such models.

Structural v.s. sufficient-stat approach in behavioral applications

- The difficulty with the structural approach is that there are often multiple positive models that can explain deviations from rationality, and each model can lead to different welfare predictions.
- The sufficient-statistic approach can be useful in such situations because welfare analysis does not require full specification of the positive model underlying observed choices.

---

[10] such as ironing and spotlighting effects in response to nonlinear price schedules

Behavioral Models
Chetty et al. (2009)

$$U(x, y) = u(x) + y$$
$$= u(x) + Z - (p + t)x \tag{48}$$

Chetty et al. dropped the assumption that the (x, y) maximizes $U(x, y)$. They took the demand function $x(p, t)$ as an empirically estimated object generated by a model unknown to the policy maker, permitting $\frac{\partial x}{\partial p} \neq \frac{\partial x}{\partial t}$.

Social welfare is given by

$$W(p, t) = \{u(x) + Z - (p + t)x\} + tx(t). \tag{49}$$

In nonoptimizing models, the envelope condition does not hold. Instead, totally differentiate the social-welfare function to obtain

$$\frac{dW}{dt} = \left[u'(x) - p \right] \frac{dx}{dt}. \tag{50}$$
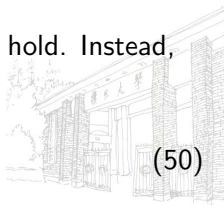
Behavioral Models
Chetty et al. (2009)

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

Given that $\frac{dx}{dt}$ can be estimated empirically, the challenge in calculating $\frac{dW}{dt}$ is the recovery of the true preferences $u'(x)$. $\Rightarrow$

Assumption 2

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

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

---

[a]This assumption requires that the agent only makes mistakes with respect to taxes, and not fully salient prices.

Behavioral Models
Chetty et al. (2009)

Let $P(x) = x^{-1}(p, 0)$ denote the agent's inverse-price-demand curve.
Assumption 2 implies that $P(x) = u'(x)$ via F.O.C. Plug into equation (50)
$\Rightarrow$

$$\frac{dW}{dt} = [P(x) - p]\frac{dx}{dt}. \tag{51}$$

To simplify implementation, Chetty et al. made the approximation that
demand $x(p, t)$ is linear in both arguments to obtain

$$\frac{dW}{dt} = \left[\frac{dp}{dx} \cdot (x(p, t) - x(p, 0))\right]\frac{dx}{dt} = \left[\frac{dp}{dx} \cdot \frac{dx}{dt}t\right]\frac{dx}{dt} = t\theta\frac{dx}{dt}, \tag{52}$$

where $\theta = \frac{dx}{dt}/\frac{dx}{dp}$ measures the degree of underreaction to the tax relative
to the price.
$\Rightarrow$ The price and tax elasticities of demand ($\frac{dx}{dp}$ and $\frac{dx}{dt}$) are together
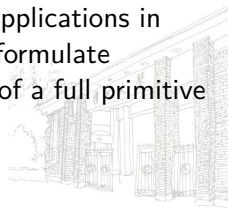sufficient stats.

Conclusion

The literature reviewed in this article has focused on identifying sufficient statistics for normative (welfare) analysis.

Sufficient statistics can also be used to answer positive (descriptive) questions.

- Ex. 1: predict the effect of a tax change on tax revenue.
- Ex. 2: capitalization effects, which shows that changes in asset prices are sufficient statistics for distributional incidence in dynamic equilibrium models.

One cannot exploit envelope conditions in most positive applications in positive analysis. However, the general concept is still to formulate answers to questions in terms of a few elasticities instead of a full primitive structure.

Application of Sufficient-Stat Approach in Other Fields

Macro

An example: whether households adhere to permanent-income hypothesis

- The structural approach: specify a dynamic model of optimization, and test whether observed consumption and savings patterns are consistent with those predicted by the model.

- The sufficient-stat counterpart: isolate one moment, such as the drop in consumption at retirement or the sensitivity of behavior to cash on hand, that is adequate to test between models.

Labor

Two examples:

- Effects of minimum wage: the optimal policy using sufficient-stat;

- Returns to schooling: examining effects on total earnings is adequate to measure the benefits of additional schooling in a model in which agents optimize.

Application of Sufficient-Stat Approach in Other Fields

Development

An example: risk-sharing arrangements

- Identifying consumption fluctuations and risk aversion is sufficient to make inferences about the welfare costs of shocks.

- $\Rightarrow$ So quasi-experimental evidence is usually enough. More generally, one may give precise answers to policy questions using estimates from randomized experiments coupled with sufficient-stat formulas derived from standard structural models.

Industrial Organization

Two challenges:

- Many questions concern discrete changes. (e.g. firm entry)

- Many questions focus on models of strategic interaction, where small changes in exogenous parameters can lead to behavioral jumps. Different techniques to apply the sufficient-stat approach are needed.

The End