Transparent Structural Estimation

Matthew Gentzkow

Fisher-Schultz Lecture

(from work w/ Isaiah Andrews & Jesse M. Shapiro)

“A hallmark of contemporary applied microeconomics is a conceptual framework that highlights specific sources of variation”

Angrist & Pischke 2010

Structural papers typically include a heuristic "identification" section…

"Loosely speaking, identification relies on three important features of our model and data…"

- Einav et al. 2013

"We now intuitively discuss the identification of [key parameters]…"

- Berry et al. 2013

"We now discuss the variation in the data that identifies each of [our key parameters]…"

- Bundorf et al. 2012

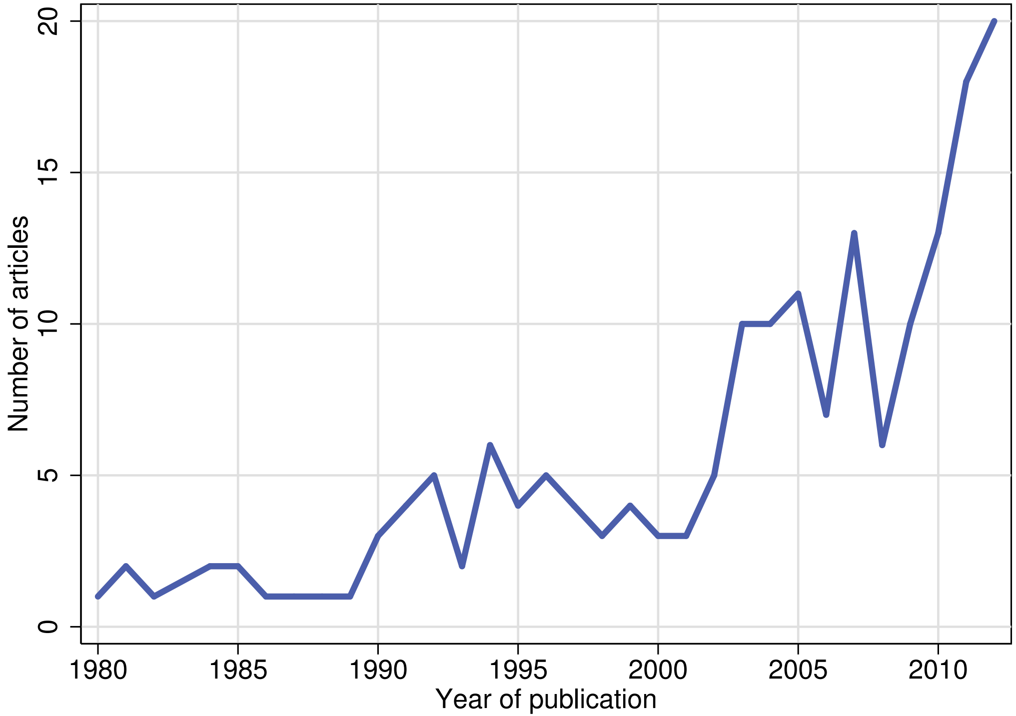
…which may contain statements like

"The **main source of identification** for θ is [moment 1]"

"The demand parameters… are **primarily identified by** [a set of moments]"

"One may think of [moment 2] as **empirically identifying** δ"

"Is Identified By"



Articles in *AER, QJE, JPE, Ecma, and ReStud*

What do these statements mean?

Why are we making them?

How can we make them more precise?

Today

New research with Isaiah Andrews and Jesse Shapiro on ways to measure what data features drive structural estimates

1. Motivating example
2. Review of identification
3. Setup
4. Measure #1: Informativeness
5. Measure #2: Sensitivity

"On the informativeness of descriptive statistics for structural estimates." Working paper, 2018.

"Measuring the sensitivity of parameter estimates to estimation moments." *QJE,* 2017.

1

Example

Valuing New Goods in a Model with Complementarity: Online Newspapers

Matthew Gentzkow
Harvard University


YOUR DISTRICT NEWS EXTRA
SCHOOLS • COMMUNITY EVENTS • HOME SALES • CRIME WATCH
The Washington Post
DISTRICT HOME EDITION
CAMPAIGN 2008
Bodies Of 4 Girls Found in SE House
Deaths Treated As Homicides
Blast Kills 6 as Troops Hunt Iraqi Insurgents
U.S. Forces Encounter Booby-Trapped House
Turning It Around
Down in polls after an Iowa loss, Hillary Clinton had no victory speech for New Hampshire. But then an unexpected thing happened: She won.
New Survey Estimates Far Fewer Iraqi Deaths


Personalize Your Post
mywashingtonpost.com
washingtonpost.com
Subscribe | Print Edition
The Washington Post
Metro Weather
69°F/20°C
P/cloudy
Traffic | DC Webcam
News
OnPolitics
Entertainment
Live Online
Camera Works
Marketplace
Jobs
Jobs FIND / POST
Cars NEW BUY / SELL
Real Estate BUY / SELL
Nation
World
Metro
Sports
Business
Technology
Style
Editorial Page
Travel
Health
Real Estate
Home & Garden
Education
Food
Outlook
Magazine
Book World
News Digest
Archives
Site Index
Corrections
Help/Feedback
Wednesday, April 30, 2003 | Updated 8:41 p.m. ET
Donald Rumsfeld talks with Maj. Gen. Robin Brimms at Basra International Airport. (AFP)
2nd U.S.-Iraqi Clash at Protest
Two more Iraqis are slain in
Mideast Peace 'Road Map' Released Amid Violence
U.S.-led plan launched shortly after approval of Palestinian prime minister. – Molly Moore and John Ward Anderson
• Palestinian Premier Denounces Terror
• Jewish Leaders Back Peace Plan
text Draft of the 'Road Map' | FAQ
Jefferson Morley: World Opinion
Plan Draws Some Skepticism
Some Palestinians and Israelis have doubts on prospects of "road map."
Greenspan Predicts a Bounce
Fed chief sees a postwar economic lift, offers mixed message on tax cuts.
SITE SEARCH:Options
News
GO
WEB SEARCH:
GO
by Google
advertising

MODEL

- Discrete choice demand model
- Choices are the set of all *bundles* of the underlying goods
- Allows for both substitutes and complements
- Allows correlated unobservables

1

DATA

- Micro data from Scarborough Research on characteristics & media consumption of 16,171 adults in Washington DC area between 2000 and 2003
- Specifically, records consumption of:
  - Washington Post
  - Washington Times
  - washingtonpost.com
- Asks what you read in last 24 hours and in last 5 days

IDENTIFICATION FROM CHOICE DATA

- After controlling for observables, readership of the post.com is *positively* correlated with readership of the Post
- Two possible reasons:
  - Post and post.com are complements
  - Unobservable consumer tastes are correlated across the two products (e.g. some consumers just have a taste for news)
- How can the data separate these given that we have no variation in prices?

IDENTIFICATION FROM CHOICE DATA

I show that there are two intuitive sources of identification…

1. **Exclusion-restrictions:** Variables that affect the utility of the post.com but not the Post

2. **Quasi-panel data:** Observe both one and five-day choices

Table 3
Linear probability model of Post consumption

	OLS	IV		
		(1)	(2)	(3)
	Dependent variable: read Post print edition last 5 days			
post.com	.0464	-.426	-.348	-.377
	(.0090)	(.106)	(.129)	(.201)
Other Internet news			.0133	.0034
			(.0181)	(.0250)
Detailed occupation controls	No	No	No	Yes
N	14313	14313	14313	10544
R-squared	.333	.208	.246	.204

Instruments: Internet access at work; fast Internet connection; use of Internet for e-mail, chatting, research/education, and work-related tasks.

RESULTS (PREVIEW)

- Accounting for observed and unobserved heterogenetiy changes the estimates from strong complements to strong substitutes
- Crowding out is moderate (removing online paper increases print readership by 1.7%)
- Optimal price is $.20/day and loss from charging zero is $9m/year
- Welfare: +$42m/year (consumers); -$20m/year (firms)
- Both introduction of online paper and pricing are close to optimal at 2004 advertising levels.

Unresolved

- How much does each source of variation drive the results? (i.e., which assumptions should a busy reader focus on?)

- How much should finding the IV evidence convincing increase confidence in the final estimates?

2

Identification

A model is **identified** if alternative values of the parameters imply different distributions of observable data

Matzkin (2013)

This is a **binary** property

Not coherent to say a parameter is *mainly*, *primarily*, or *mostly* identified by a particular moment

This is a property of a **model** not an estimator

Not coherent to say two estimators of the same model are identified differently

Perfectly correct to say θ is identified by a feature of the data even if that feature of the data doesn't enter estimation at all

In the Print-Online example…

Correct to say that both exclusion restrictions and panel data are sources of identification (i.e., model would be identified with either alone)

Not clear what it means to ask which is a more important source of identification

Not clear how IV regression should affect our confidence in the structural estimates

“What is meant by ‘identified’ is subtly different from the use of the term in econometric theory…. ‘How a parameter is identified’ refers to a more intuitive notion that can be roughly phrased as ‘What are the key features of the data… that drive [the estimates].”

Keane (2010)

“**Loosely speaking**, identification relies on…”

- Einav et al. 2013

“We now **intuitively** discuss the identification of…”

- Berry et al. 2013

“One may **casually think** of [a set of moments] as ‘empirically identifying’…”

- Crawford and Yurukoglu 2012

“[We offer a] **heuristic discussion…** Although [we treat] the different steps as separable, the… parameters are in fact jointly determined and jointly estimated.”

- Gentzkow, Shapiro and Sinkinson 2014

Nonparametric Identification

A model is **nonparametrically identified** if it is identified and it makes no assumptions about functional forms, distributions, etc. that are not grounded in economic theory (Matzkin 2013)

If a model is nonparametrically identified, there may exist a nonparametric estimator

But most structural papers that discuss nonparametric identification go on to estimate a parametric version of the model

Not clear what nonparametric identification tells us about the credibility of the actual estimates in such cases

So why discuss identification at all?

Identification analysis tells us what information in the data **could in principle be used** to answer the question

This is valuable because

- It points the way to better estimators
- It illuminates the economics of the model

Identification analysis does not tell us what information **is actually used** by any given estimator

It therefore does not tell us much about whether we should believe any given set of estimates

To answer these questions, we need a different set of analytical tools

3

Setup

Model & Estimator

- Data $D_i \in \mathcal{D}$ for $i \in 1, \ldots, n$
- Researcher assumes $D_i \sim F(\eta)$
- Quantity of interest $c(\eta)$ with true value c_0
- Estimator $\hat{c}$ of c_0

Data Features

- Low-dimensional vector of interpretable statistics $\hat{\gamma}$

- E.g.,
  - Regression coefficients
  - Treatment-control differences from experiment

Under base model $F_0^n = \times_n F(\eta_0)$

$$\sqrt{n}\begin{pmatrix}\hat{c} - c_0 \\ \hat{\gamma} - \gamma_0\end{pmatrix} \to_d N(0, \Sigma)$$

$\Sigma_{\gamma\gamma}$, $\Sigma_{c\gamma}$ are submatrices of Σ
Σ, $\Sigma_{\gamma\gamma}$ full rank and $\sigma_c^2 > 0$

Meta-Model

- A population of readers are concerned that the model $F(\eta)$ may be misspecified
- Different readers have different priors about the most relevant alternatives $\mathcal{F}$
- Assume the true value c_0 is defined independently, so we can talk about the bias of $\hat{c}$ under any $\mathcal{F}$

Research is **transparent** if readers can easily assess potential bias under the alternatives $\mathcal{F}$ they find relevant

Key idea: Easier for readers to assess how $\mathcal{F}$ affects $\hat{\gamma}$ than to assess directly how it affects $\hat{c}$

Example (Print-Online)

- $\hat{c}$: Effect of introducing post.com on Post readership
- $\hat{\gamma}$: IV regression coefficient

- Possible alternatives $\mathcal{F}$
  - Internet at Work, etc. correlated w/ taste for news (so exclusion restrictions invalid)
  - Taste for news is time varying (so panel strategy invalid)
  - Easy to see how these affect $\hat{\gamma}$; hard to see how they affect $\hat{c}$

Local Perturbations

- Following literature (e.g., Newey 1985), focus on alternatives that are local to F_0
- Implies bias from misspecification on the same order as sampling uncertainty

- Index the space of all such perturbations by
  - *Direction* $\varphi \in \Phi$
  - *Magnitude* $\mu \in \mathbb{R}$

For any direction $\varphi \in \Phi$, define a family of distributions $F_\varphi(\nu)$ for $\nu \in \mathbb{R}^+$ such that $F_\varphi(0) = F_0$

Each $F_\varphi(\nu)$ is a path passing through F_0

The **local perturbation** $\mathcal{F}_{\varphi\mu}$ with direction φ and magnitude μ is the sequence of joint distributions

$$F_\varphi^n\left(\frac{\mu}{\sqrt{n}}\right) = \times_n F_\varphi\left(\frac{\mu}{\sqrt{n}}\right)$$

Asymptotic Bias

Assuming appropriate regularity conditions, under $\mathcal{F}_{\varphi\mu}$

$$\sqrt{n}\begin{pmatrix}\hat{c}-c_0\\ \hat{\gamma}-\gamma_0\end{pmatrix}\rightarrow_d N\left(\begin{pmatrix}\mu\bar{c}_\varphi\\ \mu\bar{\gamma}_\varphi\end{pmatrix},\Sigma\right)$$

where $\mu\bar{c}_\varphi$ and $\mu\bar{\gamma}_\varphi$ are the (first-order) **asymptotic biases** of $\hat{c}$ and $\hat{\gamma}$ respectively under $\mathcal{F}_{\varphi\mu}$, and Σ is the same as in the base case.

Goal

Tools to help readers translate intuition about the bias $\bar{\gamma}$ in $\hat{\gamma}$ from various alternatives $\mathcal{F}$ into intuition about the bias $\bar{c}$ in $\hat{c}$

- Informativeness (AGS WP 2018)
- Sensitivity (AGS *QJE* 2017)

4

Informativeness

AGS (WP 2018)

Which $\hat{\gamma}$ are the most important drivers of $\hat{c}$?

What is an “important driver”?

What is an “important driver”?

Definition #1: Across alternative realizations of the data, a lot of the variation in $\hat{c}$ is explained by variation in $\hat{\gamma}$

What is an “important driver”?

Definition #1: Across alternative realizations of the data, a lot of the variation in $\hat{c}$ is explained by variation in $\hat{\gamma}$

Definition #2: Knowing that $\hat{\gamma}$ was correctly specified (i.e., $\bar{\gamma} = 0$) would significantly reduce the scope for bias in $\hat{c}$

Definition #2 (More Precise):

Let $\mathcal{B}^{\mu}$ be the set of asymptotic biases of $\hat{c}$ under local perturbations of magnitude μ

Let $\mathcal{B}_0^{\mu}$ be the set of asymptotic biases of $\hat{c}$ under local perturbations of magnitude μ for which $\bar{\gamma}_{\varphi} = 0$

Say $\hat{\gamma}$ is an important driver if $\left|\mathcal{B}_0^{\mu}\right| << |\mathcal{B}^{\mu}|$

This Paper

- New measure Δ of the **informativeness** of $\hat{\gamma}$ for $\hat{c}$

- Δ is the R^2 from a regression of $\hat{c}$ on $\hat{\gamma}$ in data drawn from their joint asymptotic distribution

- Main result:

$$\frac{B_0^{\mu}}{B^{\mu}} = \sqrt{1-\Delta}$$

- Δ can be estimated at minimal cost even in computationally challenging models

The **informativeness** of $\hat{\gamma}$ for $\hat{c}$ is

$$\Delta = \frac{\Sigma_{c\gamma}\Sigma_{\gamma\gamma}^{-1}\Sigma_{c\gamma}'}{\sigma_c^2}$$

Recall, under F_0^n :

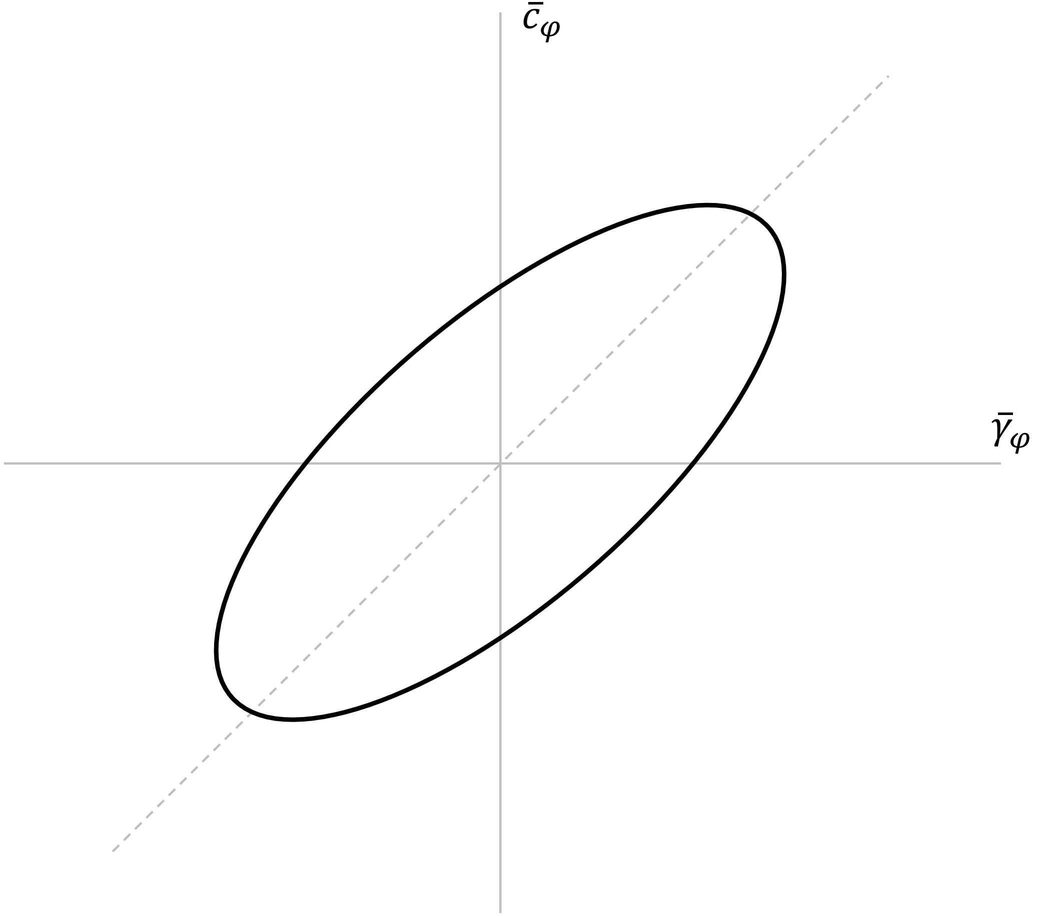
$$\sqrt{n}\begin{pmatrix}\hat{c} - c_0 \\ \hat{\gamma} - \gamma_0\end{pmatrix} \to_d N(0, \Sigma)$$

Note: Δ is unchanged under local perturbations

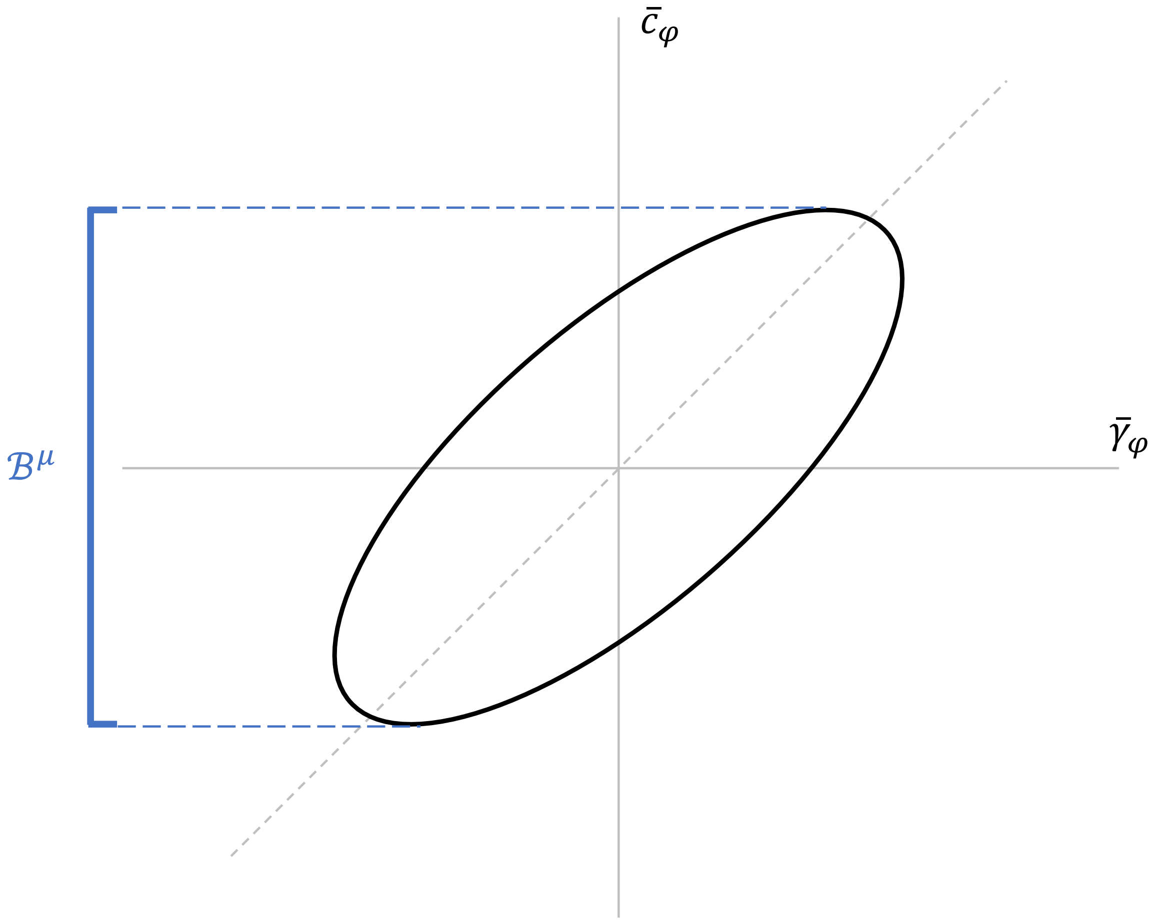
Examples

- **Minimum Distance:** $\hat{c}$ is a function of parameters estimated by minimum distance; $\hat{\gamma}$ is the vector of estimation moments; then $\Delta = 1$
- **MLE:** $\hat{c}$ is a function of parameters estimated by maximum likelihood; $\hat{\gamma}$ is a vector of coefficients from a descriptive regression; then typically $\Delta < 1$

Lemma:
Under regularity conditions, the set of all asymptotic biases $(\bar{c}_\varphi, \bar{\gamma}_\varphi)$ associated with perturbations of magnitude μ is an ellipse

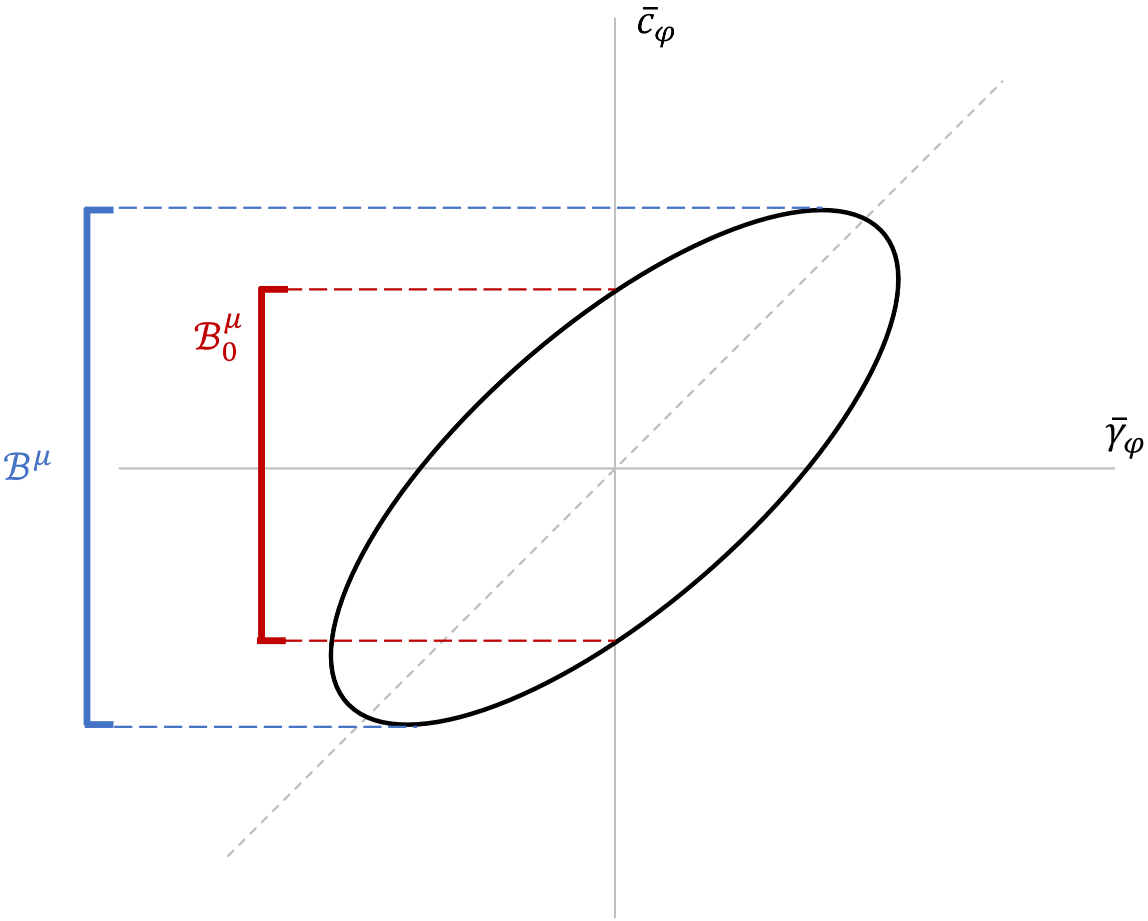


The set of asymptotic biases $\bar{c}_\varphi$ under these local perturbations



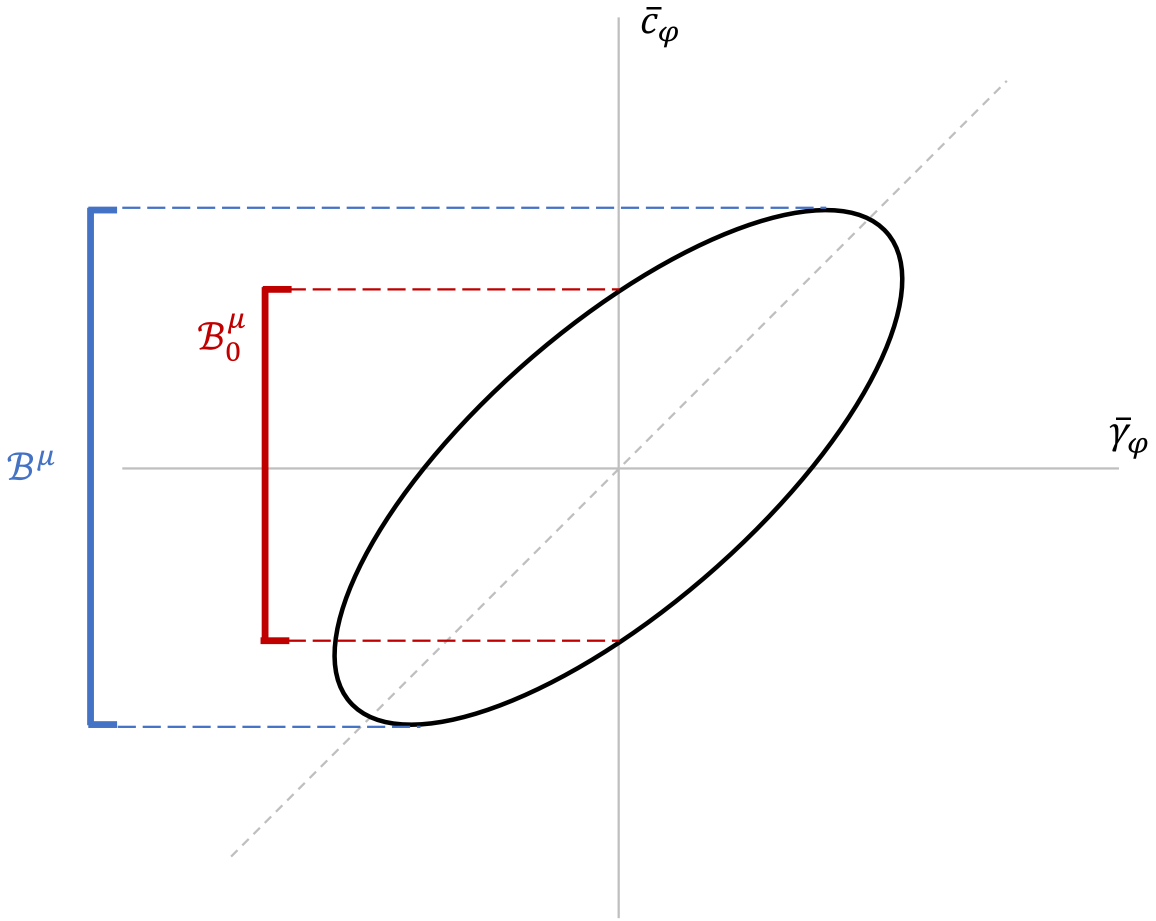
The set of asymptotic biases $\bar{c}_\varphi$ under these local perturbations

The set when we restrict to those with $\bar{\gamma}_\varphi = 0$

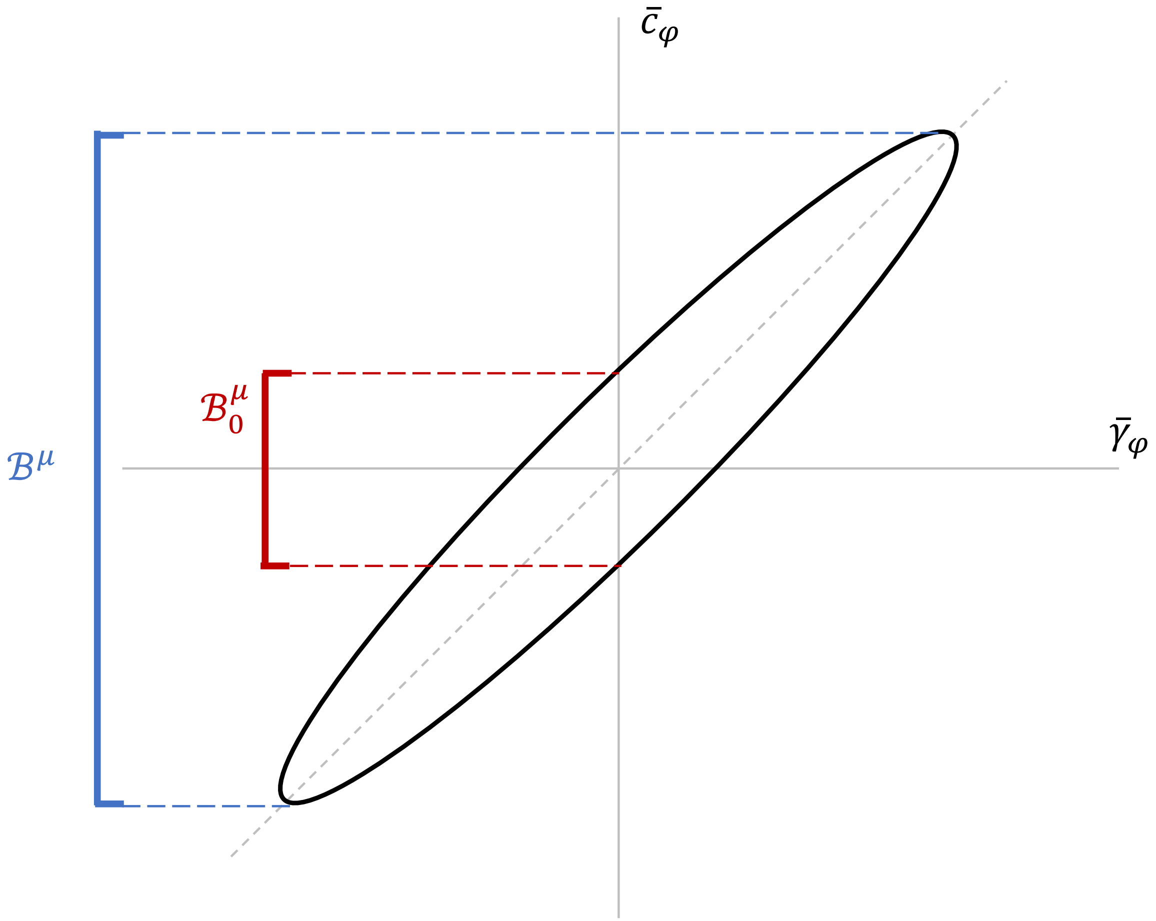


Main Result

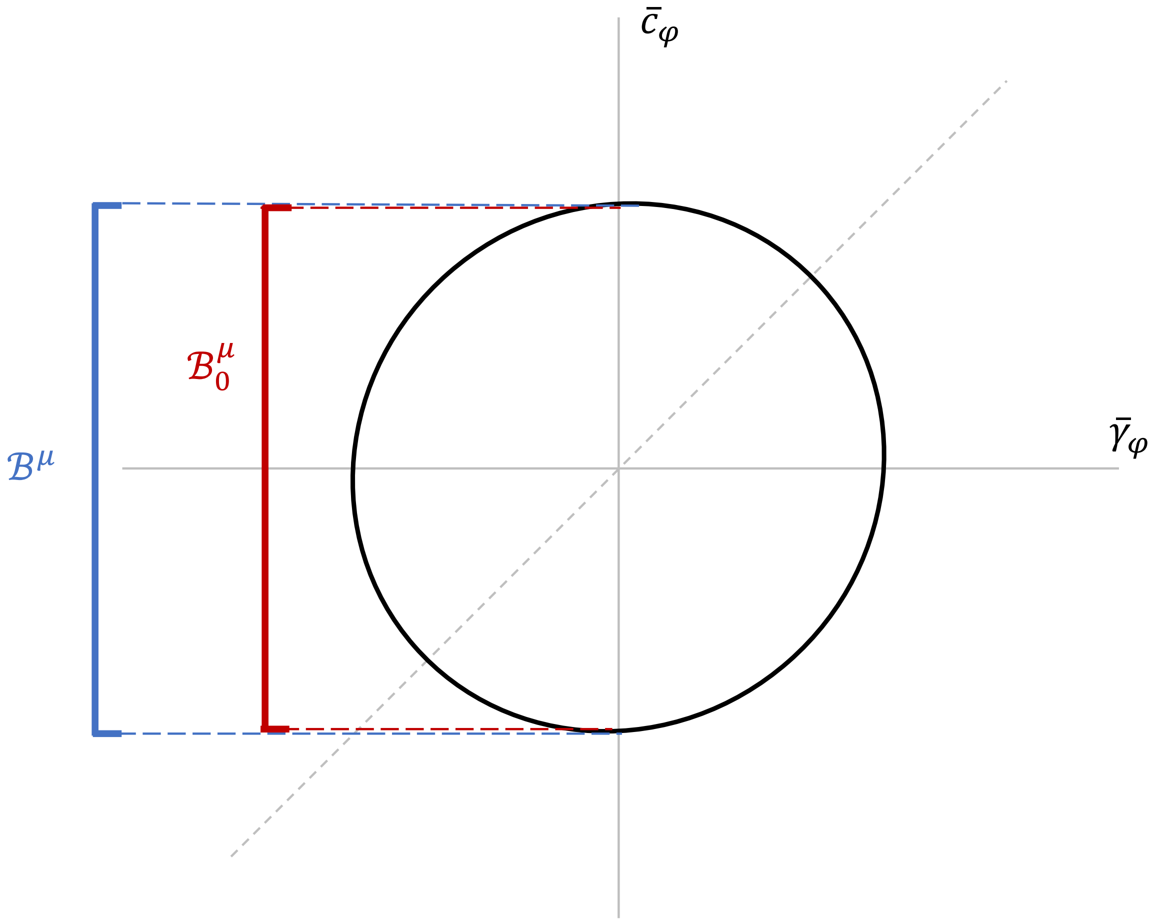
$$\frac{|\mathcal{B}_0^\mu|}{|\mathcal{B}^\mu|} = \sqrt{1-\Delta}$$



High Δ



Low Δ



An Important Subtlety

- What does it mean for $\bar{\gamma} = 0$?

- $\hat{\gamma}$ is unbiased for the true value γ_0 consistent with c_0 under the model

- Print-Online case: IV estimate must be consistent for relevant treatment effect *and* model's mapping of this treatment effect to c_0 must be correct

- That $\hat{\gamma}$ comes from a randomized experiment is *not* enough on its own

Application: Print-Online

- $\hat{c}$: Effect of introducing post.com on Post readership
- $\hat{\gamma}$:
  - IV regression coefficient (from table in paper)
  - Panel regression coefficient (crude approximation to panel variation in model)

Recall Unresolved Questions…

- How much does each source of variation drive the final estimates?

- How much should finding the IV evidence convincing increase confidence in the final estimates?

Results

Descriptive statistics $\hat{\gamma}$	Estimated informativeness $\hat{\Delta}$
All	0.635
IV coefficient	0.011
Panel coefficient	0.621

Q: How much does each source of variation drive the estimates?

A: IV variation **hardly**; Panel variation **a lot**

Q: How much should finding the IV evidence convincing increase our confidence?

A: Not much! Knowing *for sure* that the IV is valid would tighten bounds on bias by no more than 0.5%

Application: Hendren (2013)

- *Why are some groups unable to obtain insurance?*

- Use self-reports on probabilities of loss events (e.g., long-term care) along with ex post realizations to quantify private information
- Structural model maps private information to adverse selection, equilibrium outcomes, and welfare
- Estimated by MLE

Outcomes of interest $\hat{c}$

1. Fraction of focal point responses
2. Minimum pooled price ratio

Descriptive statistics $\hat{\gamma}$

1. Share in focal groups
2. Share in non-focal groups
3. Fraction in each group needing long-term care *ex post*

"The fraction of focal point responses… [is] identified from the distribution of focal points and the loss probability at each focal point"

(p. 1752)

Descriptive statistics $\hat{\gamma}$	Estimated informativeness $\hat{\Delta}$ for Fraction focal point respondents
All	0.987
Fractions in focal point groups	0.979
Fractions in non-focal point groups	0.825
Fraction in each group needing LTC	0.383

Minimum pooled price ratio will be identified by the relationship of elicited beliefs to realized losses

(p. 1751-2)

Descriptive statistics $\hat{\gamma}$	Estimated informativeness $\hat{\Delta}$ for	
	Fraction focal point respondents	Minimum pooled price ratio
All	0.987	0.700
Fractions in focal point groups	0.979	0.005
Fractions in non-focal point groups	0.825	0.018
Fraction in each group needing LTC	0.383	0.547

5

Sensitivity

AGS (*QJE* 2017)

How can we map bias $\bar{\gamma} \neq 0$ in $\hat{\gamma}$ into resulting bias $\bar{c}$ in $\hat{c}$

This Paper

- New measure Λ of the **sensitivity** of $\hat{\gamma}$ to $\hat{c}$

- Λ is the vector of coefficients from a regression of $\hat{c}$ on $\hat{\gamma}$ in data drawn from their joint asymptotic distribution

- Main result (when $\Delta = 1$):
$$\bar{c}_{\varphi} = \Lambda \bar{\gamma}_{\varphi}$$

- Λ can be estimated at minimal cost even in computationally challenging models

The **sensitivity** of $\hat{\gamma}$ to $\hat{c}$ is

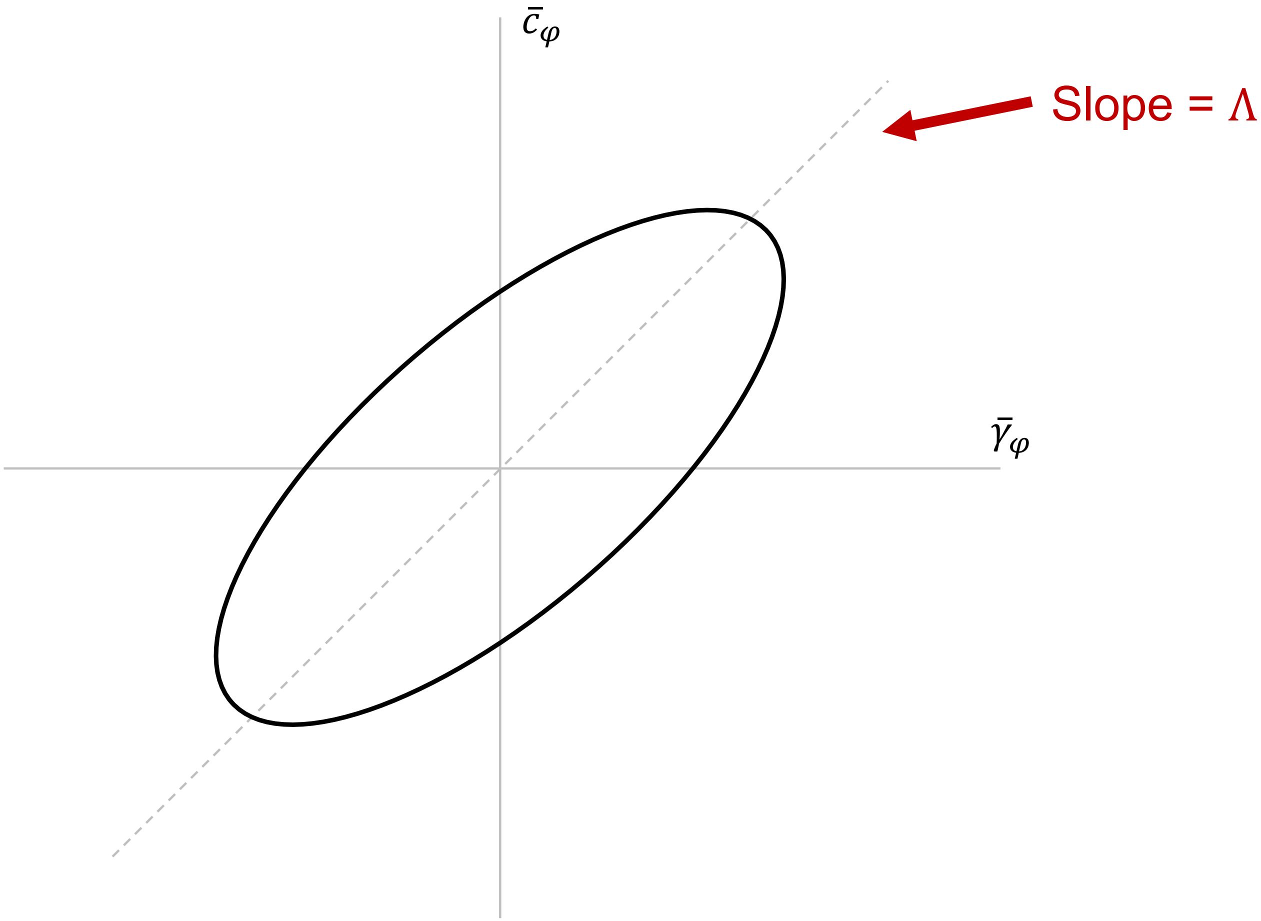
$$\Lambda = \Sigma_{c\gamma}\Sigma_{\gamma\gamma}^{-1}$$

Recall, under F_0^n :

$$\sqrt{n}\begin{pmatrix} \hat{c} - c_0 \\ \hat{\gamma} - \gamma_0 \end{pmatrix} \to_d N(0, \Sigma)$$

Note: Δ is unchanged under local perturbations

AGS (2018) extend to the case of $\Delta < 1$


$\bar{c}_\varphi$
$\bar{\gamma}_\varphi$
Slope = Λ

Sensitivity provides an analogue of the omitted variables bias formula for non-linear models

In print-online example, tells us how much a given bias in IV coefficient would affect key counterfactual

6

Conclusion

Transparent Structural Estimation

1. Show lots of descriptive evidence

- Plots of raw data
- Reduced-form regressions
- Experimental treatment effects

- Trend toward showing this kind of evidence in structural papers is a *good thing…* we should do more of it!

2. Map data features to estimator

- Show readers what data features *actually* drive key estimates
- Support these claims with evidence
- Make it easy for readers to assess the impact of misspecification they’re most worried about

- **Informativeness** and **sensitivity** provide two tools
- Please improve on these and suggest more!

3. If you discuss identification, be precise

- Statements about identification should be formal claims, ideally with rigorous proof
  - *The model is identified from the these data under the these assumptions…*

- Be clear that these are statements about what information **could in principle be used** to answer the question, not claims about what is actually used by the estimator