Recent Developments in Econometrics for Transparency and Communication

Isaiah Andrews (MIT)

Plan for this Lecture

As you likely know, the AEA has changed the structure of continuing education lectures this year

- Rather than a small number of multi-day courses after the meetings, larger number of 2-hours lectures during the meetings

In this lecture, I will discuss research that I and other have worked on, considering questions around communication and transparency in economics

- Papers of mine that I will discuss today are joint with Jesse Shapiro and in many cases Matthew Gentzkow
- These slides also draw on my joint work with them

Complete references included at end of slides

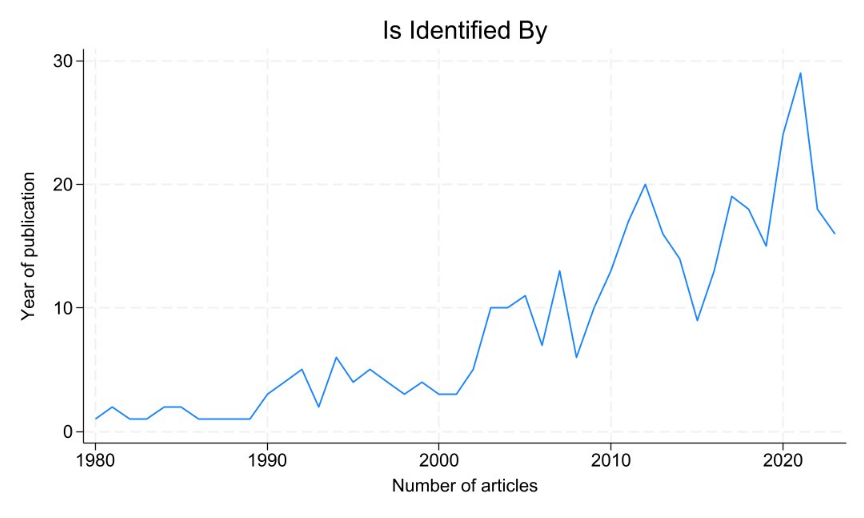
What “Identifies” Estimates

To motivate today’s lecture, start with a puzzle

Researchers often discuss what “identifies” their estimates:

- “One may casually think of [a set of moments] as ‘empirically identifying’ [a set of parameters]” (Crawford and Yurukoglu, 2012)
- “Loosely speaking, identification [of three key parameters] relies on three important features of our model and data...” (Einav et al., 2015)
- “[A spillover parameter] is identified by inventors who move across cities or across fields ...” (Moretti, 2021)
- “[A set of coefficients] is identified by variation ... within broad occupational categories within decades.” (Autor et al. 2024)

What “Identifies” Estimates



Papers using “is identified by” in AER, ECMA, QJE, JPE, and ReStud

What “Identifies” Estimates

This practice raises two questions:

1. What do we mean by these statements?
   - Don’t match the formal definition of identification
2. Why are we saying them?
   - No role in usual descriptions of “good” estimators

What “Identifies” Estimates

Formally, we say a parameter is identified if and only if its value can be recovered from the joint distribution of the data

- see e.g. Matzkin (2013)

As highlighted by Keane (2010), this mostly isn't what informal discussions of identification seem to be after

> *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]'*

What “Identifies” Estimates

This suggests a possible answer to (1):

- When we talk about what “identifies” estimates, we really mean what “drives” them in Keane (2010)’s sense

This still leaves question (2): why are we making these statements?

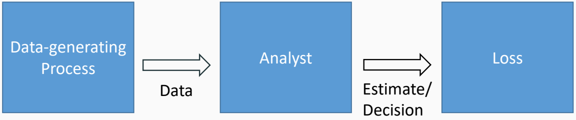
- That is, why do we care what “drives” our estimates?

Standard evaluation criteria for estimators don’t distinguish based on “drivers”

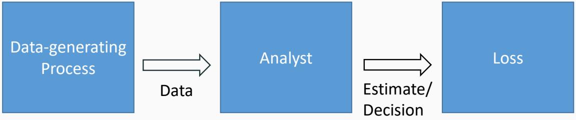
Classical Model of Statistics (Wald 1950)

- Analyst observes data $X \in \mathcal{X}$, where $X \sim F_\theta$
- Uses X to form estimate of unknown parameter $\theta \in \Theta$
- Estimate is “good” if close to true value of parameter
- Formalized by imagining a decision problem in which
  - estimate is a decision $d \in \mathcal{D}$
  - want to minimize expected loss $E_\theta \left[L(d, \theta) \right]$
- Dominant paradigm for optimal point estimation
  - e.g., $L(d, \theta) = (d - \theta)^2$ gives MSE criterion

Classical Model of Statistics (Wald 1950)



Classical Model of Statistics (Wald 1950)



- Good description of some settings: e.g. analyst works for a firm that must make a pricing decision

Classical Model of Statistics (Wald 1950)

Under classical model, what "drives" estimate is irrelevant

- If I can find an estimator $\hat{\theta}(X)$ such that

$$E_\theta\left[L\left(\hat{\theta}(X),\theta\right)\right] \leq E_\theta\left[L\left(\tilde{\theta}(X),\theta\right)\right]$$

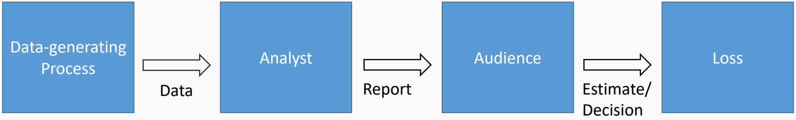
  for other candidate estimators $\tilde{\theta}(X)$, I should use $\hat{\theta}$
- e.g. in a parametric model, we should use the MLE because it beats all other reasonable estimators in large samples

Recommendation doesn't hinge on what "drives" $\hat{\theta}(X)$

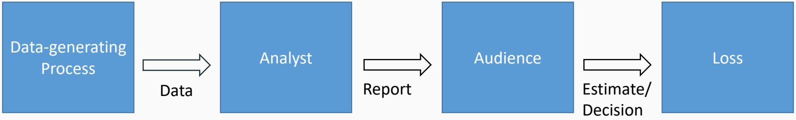
But we apparently do care what "drives" $\hat{\theta}(X)$...

- Suggests standard model is incomplete

Alternative Model of Statistics in Science



Alternative Model of Statistics in Science



- Example: Analyst reports to scientists with diverse opinions, policymakers with diverse objectives
- In this context, knowing what "drives" $\hat{\theta}$ can be valuable for audience

Plan for Today

In the rest of today's talk, I plan to:

- Formally introduce the alternative model
- Show that it gives distinct recommendations from classical model
- Develop implications for estimation and the "drivers" of estimates
- Discuss a subsequent literature building on related ideas
- Explore implications of the alternative model for another problem: uncertainty quantification

Contrasting Models of Science

Timing

Under alternative model (introduced by Andrews and Shapiro 2021)

- Analyst publicly commits to a rule $c : \mathcal{X} \to \mathcal{D}$
- Analyst observes data $X \in \mathcal{X}$, where $X \sim F_\theta$
- Analyst makes report $c(X)$ to an audience $\mathcal{A}$
- Each agent $a \in \mathcal{A}$ selects decision d and realizes loss $L(d, \theta)$

Audience

- Agents $a \in \mathcal{A}$ have different priors on θ
  - Write $\mathsf{E}_a[\cdot]$ for expectation under a's prior
  - Identify each agent with their prior, so $\mathcal{A} \subseteq \Delta(\Theta)$
- Andrews and Shapiro (2021) show can re-express heterogenous losses as heterogenous priors
  - Without loss to focus on prior heterogenity

Analyst's Goal

- Benevolent analyst wants to minimize expected loss (*risk*)
- Consider two possible notions of risk of rule c for agent a
  - *Decision risk (classical model)*

$$\mathrm{E}_a\left[L\left(c\left(X\right),\theta\right)\right],$$

    as if analyst makes decision on agent's behalf
  - *Communication risk (alternative model)*

$$\mathrm{E}_a\left[\min_{d\in\mathcal{D}}\mathrm{E}_a\left[L\left(d,\theta\right)\middle|c\left(X\right)\right]\right],$$

    as if agent makes optimal decision given report

Analyst's Goal

- In special case of squared-error loss $L(d, \theta) = (d - \theta)^2$
  - *Decision risk (classical model)*

$$\mathrm{E}_a\left[L(c(X), \theta)\right] = \mathrm{E}_a\left[(c(X) - \theta)^2\right],$$

    is mean squared error
  - *Communication risk (alternative model)*

$$\mathrm{E}_a\left[\min_{d \in \mathcal{D}} \mathrm{E}_a\left[L(d, \theta) \mid c(X)\right]\right] = \mathrm{E}_a\left[\mathrm{Var}_a(\theta \mid c(X))\right],$$

    is expected posterior variance
- Distinction irrelevant when $|\mathcal{A}| = 1$
- Distinction can matter a lot when $|\mathcal{A}| > 1$

Example

- Analyst conducts a randomized trial with a binary outcome
- Goal is to learn the success probability $\theta = (\theta_1, ..., \theta_J)$ at each of a finite set of ordered treatments $\{1, ..., J\}$
  - e.g., Probability of purchase at a set of prices
  - e.g., Probability of callback at a set of unemployment spell lengths
- Success probabilities known to be decreasing, $\theta_1 \geq \theta_2 \geq ... \geq \theta_J$
  - e.g., Demand slopes down
  - e.g., Longer unemployment spells deter employers
- Quadratic loss $L(d, \theta) = \sum_j \left(d_j - \theta_j\right)^2$

Example

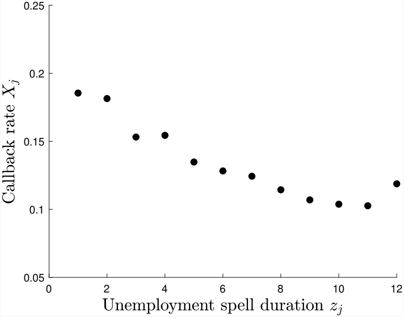
- n independent observations for each treatment
- Data $X = (X_1, ..., X_J)$ are fraction of successes for each
- Decision space $\mathcal{D} = \mathcal{X}$ rich enough to communicate full data
- Audience $\mathcal{A} = \Delta(\Theta)$ includes all possible priors
    - Everyone agrees that $\theta_j \geq \theta_{j+1}$ for all j
    - ...but may disagree about everything else

Two Rules

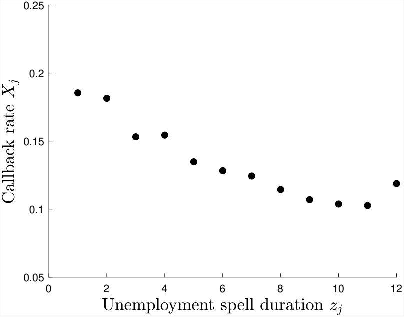
Imagine analyst can choose between two possible rules

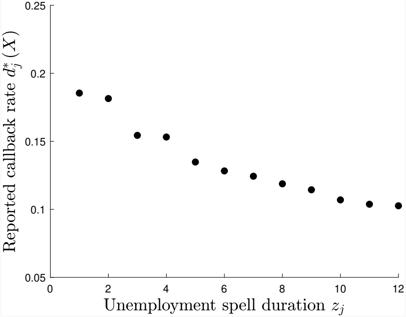
- Full data $c_j(X) = X_j$
  - Reports success fraction for each treatment j
- Rearranged data $c_j^*(X) = j$th highest element of $\{X_1, ..., X_J\}$
  - Sorts success fractions in descending order

Illustration

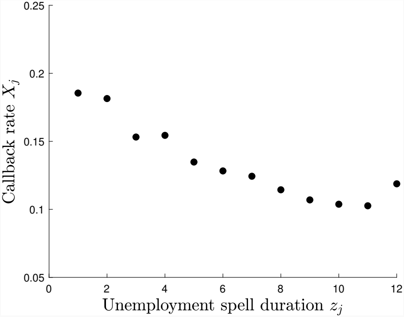


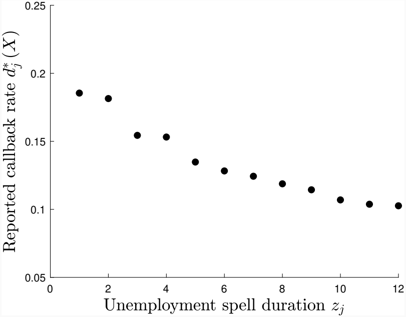
Illustration





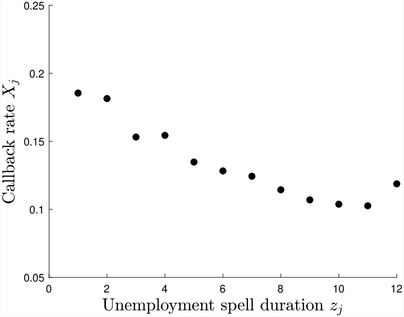
Decision Risk Perspective

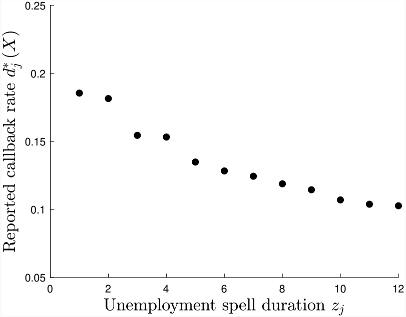




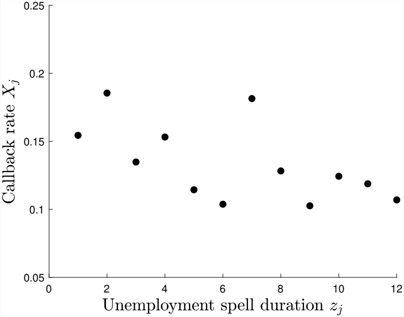
- Rearranged data c^* dominates full data c in decision risk
  - Achieves weakly lower risk for all agents, strictly lower for some
  - Intuitively, gets closer to true parameter
  - cf. Chernozhukov et al. (2009)
- Classical model would recommend c^* over c

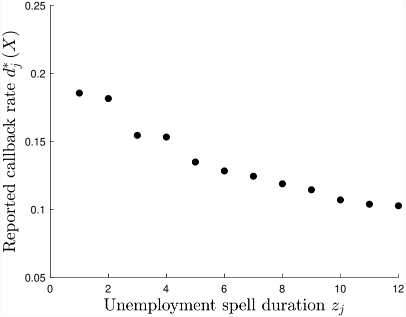
Illustration



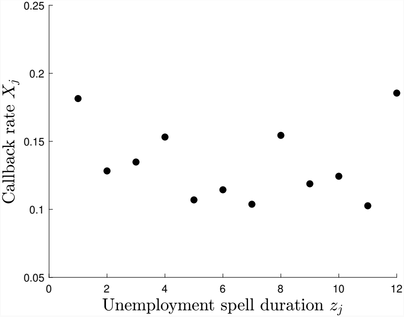


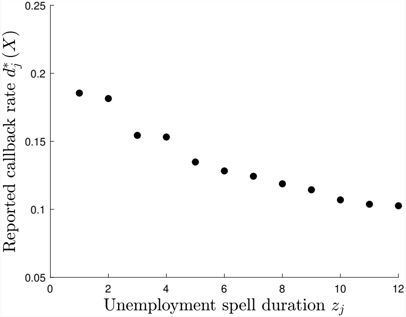
Illustration



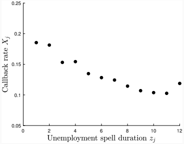


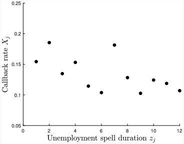
Illustration

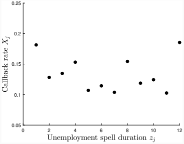


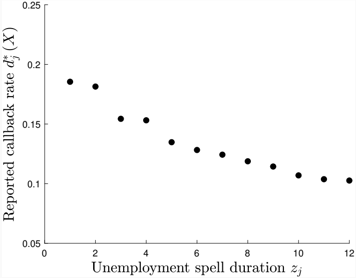


Illustration

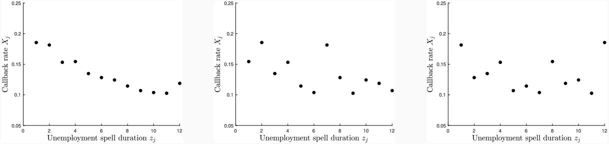


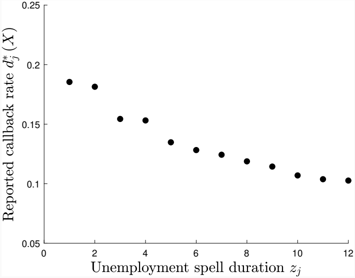






Communication Risk Perspective





- Full data c dominates rearranged data c^* in communication risk
    - Intuitively, preserves decision-relevant information in data

Conflict in Admissibility

- So far, we've shown that different models made different selections from the pair of rules $\{c, c^*\}$
- A stronger statement is true
- **Definition**: A rule is admissible (in a given notion of risk) if it is not dominated by another rule
- In this example, any rule that is admissible in decision risk is inadmissible in communication risk, and vice versa
  - Admissible decision rules sometimes discard useful information
  - Admissible communication rules sometimes recommend bad decisions
- No choice of rule resolves conflict between two notions of risk

Generalizing

- Andrews and Shapiro (2021) provide sufficient conditions for admissibility conflict in range of cases nesting example
- Shows conflict between goals of decision, which is the focus of the classical model, and communication, which seems an important part of empirical science
- *Recommendations of classical model may not achieve goals of scientific analyst who cares about communication*

Implications for Practice

- In this example, communication-optimal rules seem more in line with empirical practice
  - e.g., we're not aware of any unemployment audit studies that report only the sorted data, though many report the unsorted data
- But not always practical to report the full data X
  - One of the reasons we write papers is to condense information we present to audience
- Can we find some practical ways for analysts to reduce communication risk in such settings?

Implications for Estimation

Reports

- Let $\tau \in \mathbb{R}$ be an economic parameter of interest
  - Structural parameter, causal effect, welfare under counterfactual policy, etc.
  - $\tau = \tau(\theta)$. Suppress dependence on θ for simplicity
- Empirical research might report
  - Estimate $\hat{\tau}$ of τ obtained under some benchmark assumptions
  - Other statistics $\hat{s}$
    - Standard errors
    - Descriptive/summary statistics that might be informative about τ
    - "Placebo"/falsification tests that might speak to plausibility of researcher's assumptions under which τ is estimated
    - **Auxiliary statistics** that might improve interpretability of $\hat{\tau}$
- "Drivers" of estimates can be cast in last category

Linear IV Example

- Data $X = \{(Y_i, W_i, Z_i)\}_{i=1}^n$, all mean-zero scalars
- Economic parameter of interest is τ with

$$Y_i = W_i\tau + \varepsilon_i$$

- Under instrument exclusion ($E\left[Z_i\varepsilon_i\right] = 0$) and relevance ($E\left[Z_i W_i\right] \neq 0$), IV estimator

$$\hat{\tau} = \frac{\sum_i Z_i Y_i}{\sum_i Z_i W_i}$$

  is consistent, $\hat{\tau} \xrightarrow{p} \tau$

Linear IV Example

- But what if exclusion restriction fails, so $E[Z_i\varepsilon_i] \neq 0$?
- In that case,

$$\hat{\tau} \xrightarrow{p} \tau + \frac{E[Z_i\varepsilon_i]}{E[Z_iW_i]} = \tau + sb$$

  for

$$s = \frac{E\left[Z_i^2\right]}{E[Z_iW_i]}, \ b = \frac{E[Z_i\varepsilon_i]}{E\left[Z_i^2\right]}$$

  cf. Conley et al. (2012), who develop this argument for general IV setting
- s describes how endogenity of instrument translates to bias in point estimate
  - One version of what "drives" estimate

Audience

- Suppose each agent a has
  - continuous joint prior on (τ, s)
  - $\Pr_a\{b = b_a\} = 1$, $b_a \neq 0$
- Then a believes that $\hat{\tau}$ is inconsistent for τ , with limiting bias sb_a

$$\hat{\tau} \xrightarrow{p} \tau + sb_a$$

- For $\hat{s}$ a consistent estimate of s, the report of $(\hat{\tau}, \hat{s})$ allows agent a to construct bias-corrected estimator of τ

$$\hat{\tau}_a = \hat{\tau} - \hat{s}b_a,$$

  where a thinks that $\hat{\tau}_a$ is consistent, $\hat{\tau}_a \rightarrow_p \tau$

What “Drives” Estimates

- Hence, if report
$$c(X) = (\hat{\tau}, \hat{s}),$$
each agent a can construct an estimate they believe is consistent
- By contrast, if instead report
$$c'(X) = \hat{\tau},$$
then agents cannot construct a consistent estimates in general
- Hence, clear benefit of reporting what “drives” estimates: allows agents worried about instrument to form consistent estimates
    - Without needing access to the underlying data

Nunn and Wantchekon (2011)

- What are long-run effects of slave trade on trust levels in Africa?
  - Nunn and Wantchekon (2011) study using survey data on trust levels for respondents in different ethnic groups
  - Among other strategies, instrument exposure to slave trade with group's distance from coast during slave trade period
- Main result: past exposure to slave trade lowers trust levels today
- Potential concern: historical proximity to coast could be correlated with other variables that affect trust,
  - e.g. historical ease of travel
  - In which case IV estimates will be inconsistent

Nunn and Wantchekon (2011)

- Evaluate using Conley et al. (2012) approach
- Sensitivity $\hat{s}$ is positive, since proximity to coast correlates with more exposure to slave trade
  - Hence to overturn positive estimate $\hat{\tau}$, would need $b > 0$ so $sb > 0$
  - i.e. need that proximity to coast correlates positively with omitted variables that determine trust
- Benchmark b in placebo exercise, looking at relationship of trust and distance to coast outside Africa
  - Estimates uniformly small
  - Mostly suggest $b < 0$, so higher trust near coast

Nunn and Wantchekon (2011)

- Readers convinced by placebo exercise would think *b* small in magnitude and probably < 0
- Hence, such readers should think Nunn and Wantchekon (2011) result is “real”
  - Though could be an under-estimate
- With more precise beliefs, could bias-correct published estimate

More General Setting

This idea generalizes beyond linear IV

- Linear IV is a special case of

$$\hat{\tau} = \arg\min_{\tau} \hat{g}(\tau)' \hat{\Omega} \hat{g}(\tau)$$

  where $\hat{g}(\tau)$ is a moment condition and $\hat{\Omega}$ is a weight matrix
  - In linear IV, $\hat{g}(\tau) = \frac{1}{n}\sum_i (Y_i - \tau W_i) Z_i$
  - Many estimators can be cast in this form
- Usual identifying assumption that moment condition is zero in population implies

$$\hat{g}(\tau) \xrightarrow{p} 0$$

Asymptotic Behavior

- Suppose that we are concerned that identifying assumptions may be violated in that
$$\hat{g}(\tau) \xrightarrow{p} b$$
for b nonzero
- Then under regularity conditions and b small, Andrews, Gentzkow, and Shapiro (2017) show that
$$\hat{\tau} \overset{p}{\dashrightarrow} \tau + s'b$$
for
$$s = -\left(\frac{\partial}{\partial\tau} g(\tau)' \,\Omega \frac{\partial}{\partial\tau} g(\tau)\right)^{-1} \frac{\partial}{\partial\tau} g(\tau)' \,\Omega$$
a *sensitivity* vector
- Hence, s again describes how violations of our identifying assumptions translate to bias in our estimates
  - One answer to what “drives” our estimates

Asymptotic Behavior

- As in IV example, suppose agent a thinks $Pr_a\{b = b_a\} = 1$. Then for $\hat{s}$ a consistent estimator of s, a also believes that

$$\hat{\tau}_a = \hat{\tau} - \hat{s}' b_a$$

  is consistent for τ

  - Provided bias b_a is small, so can rely on linear approximation

- Hence, reporting

$$c(X) = (\hat{\tau}, \hat{s})$$

  allows each agent to construct an estimate they believe is consistent

Application

- Reporting $\hat{s}$ can make estimate $\hat{\tau}$ more interpretable to audience members concerned about identifying assumptions
- Computing $\hat{s}$ is easy in many important contexts
  - Already compute $\frac{\partial}{\partial \tau}\hat{g}(\hat{\tau})$ and $\hat{\Omega}$ as part of standard error calculation
- Apply in several cases
  - Automobile demand (Berry et al. 1995)
  - Lifecycle consumption (Gourinchas and Parker 2002)
  - **Charitable giving (DellaVigna et al. 2012)**

DellaVigna et al. (2012)

THE
QUARTERLY JOURNAL
OF ECONOMICS

Vol. 127 February 2012 Issue 1

TESTING FOR ALTRUISM AND SOCIAL PRESSURE IN CHARITABLE GIVING*

STEFANO DELLAVIGNA
JOHN A. LIST
ULRIKE MALMENDIER

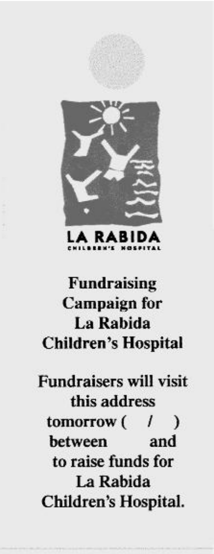
What are the welfare effects of door to door charitable solicitations?

- Provide opportunity to contribute
- But social pressure may impose costs

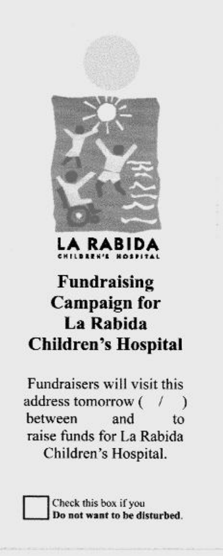
Experimental Treatments

No flyer

Flyer



Opt-out



Model

Two-period game between solicitor and household

Household preferences

- Utility from giving
- Disutility from social pressure
    - Indexed by parameter τ
    - Decreases linearly in donation amount until threshold
    - Social pressure leads to bunching of donation amounts at threshold
    - Threshold assumed to be $10 (median donation amount)

Estimation

Minimum distance estimator targeting observed shares

- opening door
- opting out
- donating various amounts

In different treatment arms

Economic Intuition

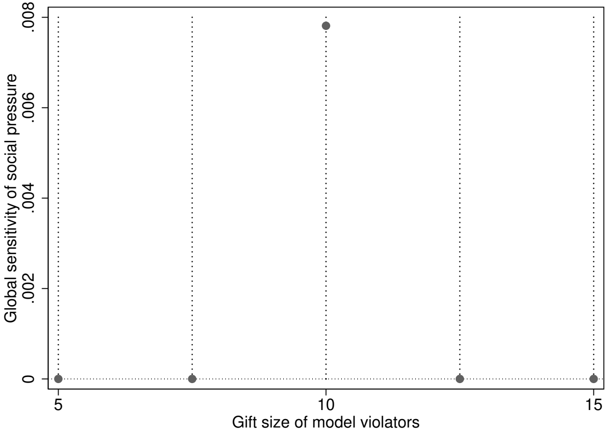
"Finally, the social pressure [τ] is identified from two main sources of variation: home presence in the flyer treatment... and the distribution of small giving (the higher the social pressure, the more likely is small giving and in particular bunching at [the threshold])."

- DellaVigna et al. (2012, p. 38)

Possible Form of Misspecification

- Suppose that some households have another reason for wanting to give exactly the threshold amount of \$10
  - e.g., they have a particular denomination of cash at home
- Audience might wonder what effect would have on the estimate $\hat{\tau}$?
- If authors had anticipated concern, could conduct a simulation or sensitivity analysis that allows for such households to be present and compute the sensitivity of $\hat{\tau}$...

Bias in Estimated Social Pressure

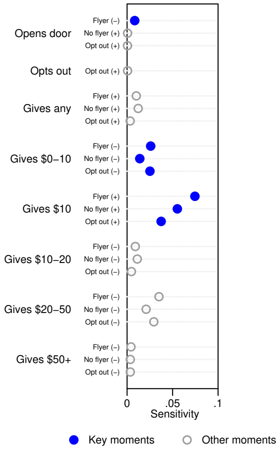


- Assumes 1% of households have exogenous giving amount
- Baseline estimate is $\hat{\tau} = 0.14$, $SE\,(\hat{\tau}) = 0.08$

Value of Reporting Sensitivity

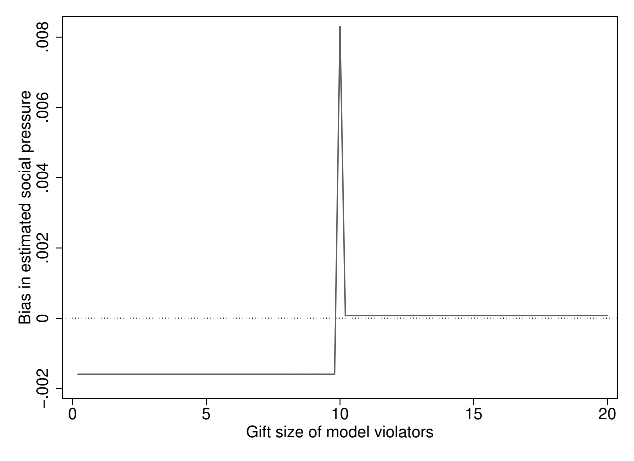
- Authors cannot anticipate and account for every possible form of misspecification that a reader may worry about
- But feasible to report estimate $\hat{s}$ of sensitivity of $\hat{\tau}$ to violations b of identifying assumptions
- Readers can use $\hat{s}$ to form own judgments about implications of these violations for parameter estimates

Sensitivity $\hat{s}$ of Estimated Social Pressure $\hat{\tau}$



Allows agents to form estimates $\hat{s}' b_a$ of bias in $\hat{\tau}$

Approximate Bias of Estimated Social Pressure



Suggests good performance of asymptotic approximation

Upshot on Sensitivity

- Sensitivity provides a formal measure for extent to which different moments “drive” estimates
  - Easy to compute with byproducts of usual standard error calculation
- From communication perspective, valuable because allows readers to judge for themselves the effect of plausible forms of misspecification
  - Similar to omitted variable bias formula for linear models

Recent Applications of Sensitivity

Sensitivity estimate $\hat{s}$ have been reported in a variety of recent papers estimating nonlinear models, e.g.

- Ottonello and Winberry (2020), who study the transmission of monetary policy
- Chen et al. (2021), who examine firm responses to an R&D subsidy program
- Balke and Lamadon (2022), who use an equilibrium model to study worker-firm risk-sharing
- Bernard et al. (2022), who study the role of firm size heterogeneity in production networks
- Kreindler (2024), who estimates the effects of counterfactual congestion-pricing policies

Recent Theoretical Results Related to Sensitivity

Recent Work on Sensitivity

There's been a variety of other work on topics related to sensitivity:

- Generalizations of sensitivity to other settings (Andrews Gentzkow Shapiro, 2020)
- Interpretation of shift-share IV (Goldsmith-Pinkham et al. 2020)
- Informativeness of estimation moments (Honoré et al., 2020)
- Sensitivity to calibrated parameters (Jørgensen, 2023)
- Misspecification-robust estimation and inference (various authors)

I'll next cover each of these in more detail

Sensitivity to Descriptive Statistics

- Andrews, Gentzkow, and Shapiro (2020) considers the case where hard to articulate beliefs in terms of $\hat{g}(\tau)$
    - Because moment conditions not economically intuitive
    - e.g., MLE
- Show how to adapt approach to this setting
    - Exploit asymptotic relationship of $\hat{\tau}$ to intuitive descriptive statistics
- Applications to
    - Gentzkow (2007)
    - Attanasio et al. (2012)
    - Hendren (2013)

Interpretation of Shift-Share IV

Many recent applied papers in economics use shift-share instruments

- Also called Bartik instruments, after Bartik (1991)
- Tabulation by Goldsmith-Pinkham (2024) suggests that 1/8th of IV strategies used in NBER working papers since 2013 are shift-share

To fix ideas, suppose we are interested in effects of a treatment W_i (e.g. employment growth) on an outcome Y_i (e.g. wage growth)

- Suppose further that i represent metro areas
- Have location-specific industry shares Z_{ik} and employment growth rates G_{ik}, so

$$W_i = \sum_k Z_{ik} G_{ik}$$

Interpretation of Shift-Share IV

- We'd usually worry that employment growth rates are endogenous
- Shift-share idea: suppose instrument W_i with

$$Z_i = \sum_k Z_{ik} G_k$$

  where G_k is national average industry growth
  - Instrument only "keeps" variation from location-specific industry mix and national industry growth rates
- Can be justified in two ways:
  - Treat industry growth rates as exogenous (Adao et al. 2019, Borusyak et al. 2022)
  - **Treat industry shares as exogenous** (Goldsmith-Pinkham et al, 2020)

Interpretation of Shift-Share IV

- Goldsmith-Pinkham et al. (2020) show that can write shift-share IV estimate as linear combination of IV coefficients using one industry as a time
$$\hat{\tau} = \sum_k \hat{\alpha}_k \hat{\tau}_k$$
where $\sum_k \hat{\alpha}_k = 1$, though may have $\hat{\alpha}_k < 0$
- As discussed in Goldsmith-Pinkham et al. (2020), $\hat{\alpha}$ related to $\hat{s}$
    - $\hat{\alpha}_k$ measures sensitivity of $\hat{\tau}$ to $\hat{\tau}_k$
- Recommend reporting which industries have large $\hat{\alpha}_k$
    - Clarifies which industries "drive" results
    - Also reporting values of $\hat{\alpha}_k$ would allow bias correction as in last section

Informativeness of Estimation Moments

Honoré et al. (2020) take a different perspective on what it means for a moment to "drive" conclusions

- Examine impact of different moments on the variance of estimates
  - Under their definition, an "important" moment is one that matters a lot for the variance of $\hat{\tau}$
- Propose a variety of sensitivity measures,
  - Measuring e.g. the change in $Var(\hat{\tau})$ if we slightly increase the variance of one moment
- Under regularity conditions, can again be estimated based on byproducts of usual variance calculations

Informativeness of Estimation Moments

Honoré et al. (2020) sensitivity measure can again be understood through lens of communication model

- Consider an agent with $Var_a(b) > 0$
  - So thinks there may be bias, but uncertain about value
- For this agent, misspecification variance acts like an increase in the variance of the moments
  - Hence, $Var_a(\hat{\tau} - \tau)$ is equal to $Var(\hat{\tau})$ plus $Var_a(b)$ times Honoré et al. (2020) sensitivity measure

Hence, this measure describes how agents's assessment of estimate changes with their uncertainty about misspecification

Sensitivity to Calibrated Parameters

Jørgenson (2023) addresses another possible audience concern: calibrated parameters

- In many settings, researchers calibrate some parameters to values taken from literature
- Estimate remaining parameters on their data
- Yields estimate $\hat{\tau}_\beta$ based on calibrated parameters β

A reader might disagree with some calibration choices

- e.g. for most economically interesting parameters, literature contains a range of credible estimates
- Re-estimating model for all plausibly interesting choices is extremely computationally heavy

Sensitivity to Calibrated Parameters

Jørgenson (2023) proposes a simple approximation to $\frac{\partial}{\partial \beta}\hat{\tau}_\beta$

- Can again be computed without re-estimating model, and mechanics are similar to $\hat{s}$

Can again interpret through communication model

- If consider an audience of agents, each of whom has a preferred β_a, and approximation error small, then each agent a thinks

$$\hat{\tau}_\beta + \frac{\partial}{\partial \beta}\hat{\tau}_\beta \left(\beta_a - \beta\right)$$

  is a consistent estimator
- Hence, reporting $c\left(X\right) = \left(\hat{\tau}, \frac{\partial}{\partial \beta}\hat{\tau}_\beta\right)$ fully resolves uncertainty stemming from choice of calibrated parameters

Misspecification-Robust Procedures

Sensitivity measures are one response to misspecification concerns

- Give audience ingredients to compute estimates they may prefer

Another response: change the estimator/confidence interval

- If there's an alternative model everyone likes better, this seems the obviously correct answer
    - If we know the way in which our model is wrong, we should fix it
- However, what if different audience members are concerned about different forms of misspecification?
- Or we all think the model is wrong, but are not sure in which way?

Misspecification-Robust Procedures

- Armstrong and Kolesar (2021), Bonhomme and Weidner (2022), Christensen and Connault (2023) address variants of this case
  - Considering point estimation, confidence intervals, or both
- These papers specify a metric for measuring the degree of misspecification, and an upper bound on severity
  - e.g. in Armstrong and Kolesar (2021), could measure violation of moments by

    $$\|b\|_1 = \sum |b_j|\,,\ \|b\|_2 = \sqrt{\sum b_j^2},\ \text{or}\ \|b\|_\infty = \max_j |b_j|$$

    and impose that less than some level $M \in \mathbb{R}_+$
- Derive improved estimators and confidence intervals which account for possibility of misspecification

Misspecification-Robust Procedures

- Robust procedures actively address misspecification concerns
  - But requires choice of metric and bound
- If unsure of bound, papers suggest reporting results for a range
  - Audience members with a view on likely severity of misspecification can focus on results relevant for them
- Unlike sensitivity, doesn't facilitate bias-correction when audience members have precise beliefs
- "Right" approach depends nature of misspecification concern
  - Sensitivity may be preferable in cases where audience knows how they would like to adjust estimation (e.g. failure of exclusion)
  - Bounds/robust estimation preferable when audience thinks assumptions wrong, but unsure of direction (e.g. relaxing logit)

Communication and Uncertainty Quantification

Communication and Uncertainty Quantification

Communication issues can be important even when we're not worried about model specification

To illustrate, close by discussing another problem where communication matters: uncertainty quantification

Communication and Uncertainty Quantification

- We often summarize our conclusions from data using point estimates and standard errors,

$$c(X) = (\hat{\tau}, \hat{\sigma})$$

- Usual justification: asymptotic arguments, e.g. central limit theorem, which tell us that

$$\frac{\hat{\tau} - \tau}{\hat{\sigma}} \approx^d N(0, 1)$$

  in large samples, under regularity conditions
- $(\hat{\tau}, \hat{\sigma})$ then inform decisions by audience
  - e.g. by policymakers, firms, and households
  - Decisions often depend not just on point estimate $\hat{\tau}$, but also on standard error $\hat{\sigma}$

Communication and Uncertainty Quantification

Usual practice also justified from communication perspective

- Bernstein-von Mises (BvM) theorem states that for $\hat{\tau}$ an efficient estimator, $\hat{\sigma}$ its standard error, and very wide class of priors π

$$\pi(\tau|X) \approx N\left(\hat{\tau}, \hat{\sigma}^2\right)$$

  in large samples, under regularity conditions
    - Originally proved for parametric models and $\hat{\tau}$ MLE
    - More recently has been extended to semiparametric settings, e.g. GMM
- Implication: reporting $c(X) = (\hat{\tau}, \hat{\sigma})$ is “nearly” a sufficient statistic

Failure of Normal Approximation

There are lots of cases where we know normal approximation fails:

- Failure or near-failure of point identification
  (e.g. Staiger and Stock 1997)
- Parameters close to the boundary of the parameter space
  (e.g. D. Andrews 1999)
- Heavy-tailed data (e.g. Cont 2001)
- Non-differentiability of model or parameter
  (e.g. Hirano and Porter 2012)
- Highly nonlinear models or target parameters
  (e.g. Andrews and Mikusheva 2016)

Failure of Normal Approximation

In such cases, BvM theorem tends to fail as well

- Full data beliefs $\pi(\tau|X)$ may be very different from $N(\hat{\tau}, \hat{\sigma}^2)$
  - Usual report may mislead us about degree of uncertainty
- Econometrics literature has developed solutions in some cases
  - e.g. robust confidence intervals
  - But are problem-specific, and can be computationally impractical
- Communication perspective offers another route

Comparison of Distributions

- If we knew audience had prior π, could compute

$$d\left(\pi\left(\tau|X\right), N\left(\hat{\tau}, \hat{\sigma}^2\right)\right)$$

  for some distance measure d
- If large, indicates BvM approximation is poor for this prior
  - Suggests that just reporting $c\left(X\right) = \left(\hat{\tau}, \hat{\sigma}\right)$ may lose useful information relative to reporting $c'\left(X\right) = \pi\left(\tau|X\right)$ or X
- By contrast, if distance small then usual report is fine

Comparison of Distributions

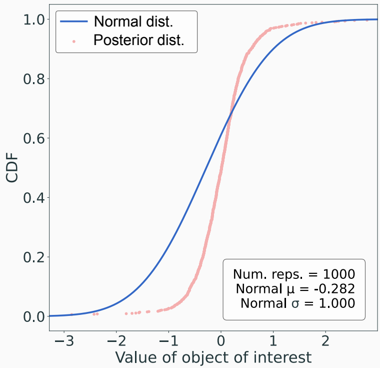
- Andrews and Shapiro (2024) makes this argument precise
- Under a class of priors and loss functions, show worst-case increase in communication risk from reporting

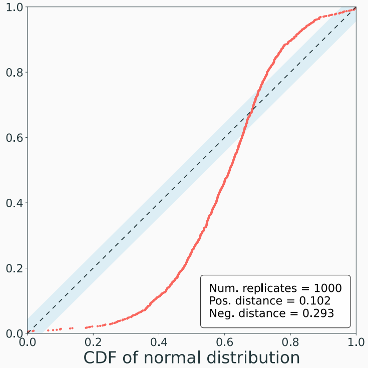
$$c(X) = (\hat{\tau}, \hat{\sigma})$$

rather than full data is bounded by (a multiple of) a particular distance d between $N\left(\hat{\tau}, \hat{\sigma}^2\right)$ and $\pi(\tau|X)$

Comparison of Distributions

Distance is sum of largest negative and positive distances between distribution function of $N\left(\hat{\tau}, \hat{\sigma}^2\right)$ and that of $\pi\left(\tau|X\right)$





Call this signed Kolmogorov distance

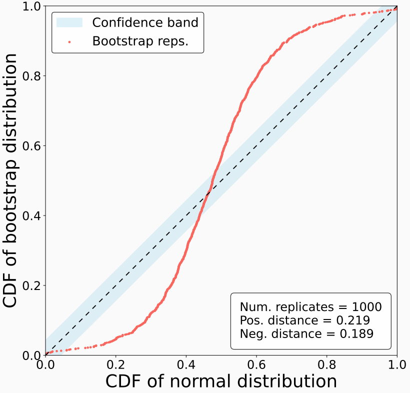
Comparison of Distributions

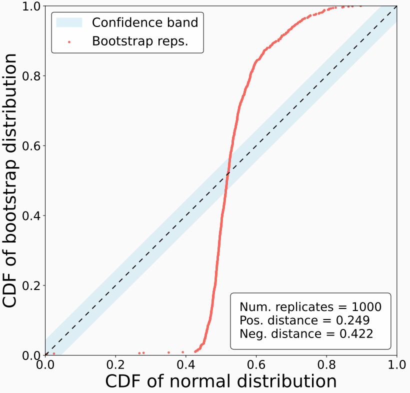
- Suggest (Bayes) bootstrap distribution as default posterior $\pi\left(\tau|X\right)$
  - Similar to nonparametric bootstrap, but randomly reweights observed data according to a different distribution
  - Corresponds to posterior under a particular, nonparametric, uninformative prior on data distribution
- If $N\left(\hat{\tau}, \hat{\sigma}^2\right)$ and bootstrap distribution very different, implies that losing substantial information by reporting $c\left(X\right) = \left(\hat{\tau}, \hat{\sigma}\right)$
  - For an agent with particular nonparametric prior
  - Simple improvement: report bootstrap distribution
- Explore quality of normal approximation in papers from the AER

AER Sample

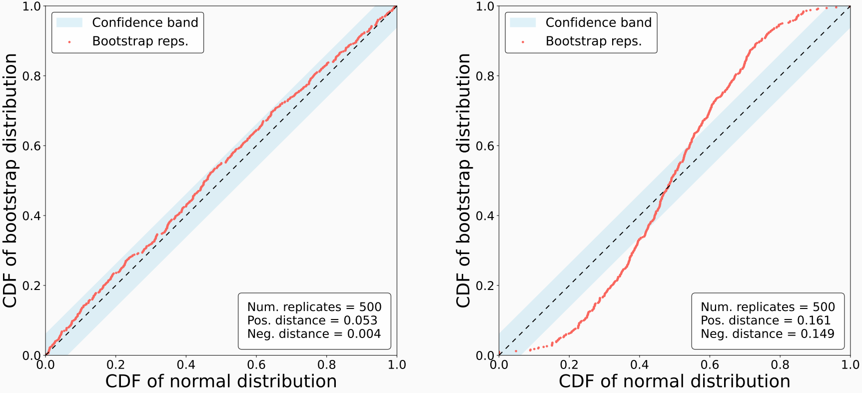
- Sample: *AER* 2021 empirical papers that use the bootstrap for some object of interest
  - Quantitative or qualitative description appears in abstract or introduction
- Wide range of topics, models, and target parameters
  - e.g., parameters describing technology, welfare calculations from a structural model, transformations of regression coefficients
  - When possible, reproduced bootstrap replicates using authors' code
  - Otherwise, contact authors for replicates
- Obtained replicates for all but one paper
  - 81 objects of interest for 14 papers
- Next, P-P plots for objects of interest from 3 of 14 papers
  - 2 OOIs from each paper with smallest and largest distances

PP Plots for $N\left(\hat{\theta}, \widehat{\sigma}^2\right)$ and Bootstrap Distribution

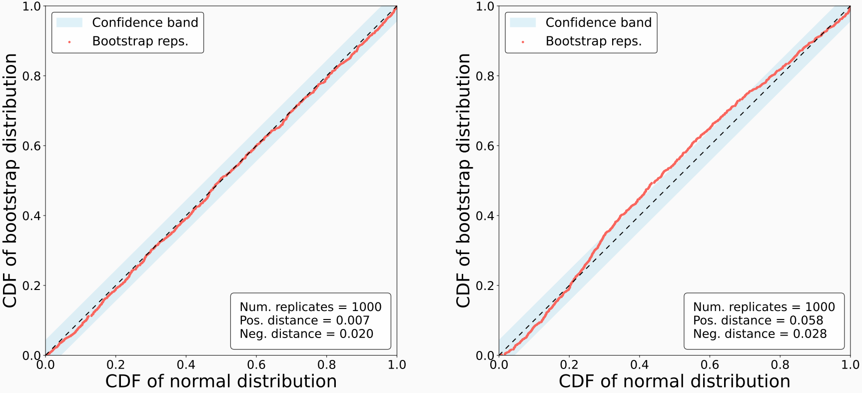




PP Plots for $N\left(\hat{\theta}, \widehat{\sigma}^2\right)$ and Bootstrap Distribution



PP Plots for $N\left(\hat{\theta}, \widehat{\sigma}^2\right)$ and Bootstrap Distribution



Takeaways for Uncertainty Quantification

- Quality of normal approximation varies greatly across papers, and across objects of interest within a given paper
- When distance between normal and bootstrap is large, only reporting $c(X) = \left(\hat{\theta}, \widehat{\sigma}\right)$ gives a misleading picture of uncertainty
  - Bad from a communication perspective
- If we're happy with the priors leading to bootstrap, easy fix: report bootstrap distribution
- If we want to guarantee e.g. coverage, more work is needed
  - Must diagnose source of problems and compute appropriate robust confidence intervals
  - And deal with pre-testing concerns

Conclusions

Summary

- Focusing on communication rather than decision-making changes understanding of the goals of empirical scientist
- Emphasizes allowing audience to make informed judgments rather than picking the single best decision/estimator for all
- Suggests practical approaches to improving transparency of empirical research in some common settings

Thank you!

References

- Crawford, G.S., and A. Yurukoglu. 2012. The welfare effects of bundling in multichannel television markets. American Economic Review. 102(2): 643-685.
- Einav, L., A. Finkelstein, and P. Schrimpf. 2015. The response of drug expenditures to non-linear contract design: Evidence from Medicare Part D. Quarterly Journal of Economics. 130(2): 841–899.
- Moretti, E. 2021. The Effect of High-Tech Clusters on the Productivity of Top Inventors. American Economic Review. 111(10). 3328–75.
- Autor, D., C. Chin, A. Salomons and B. Seegmiller. 2024. New Frontiers: The Origins and Content of New Work, 1940–2018. The Quarterly Journal of Economics. 39(3). 1399–1465.

References

- Matzkin, R.L. 2013. Nonparametric identification in structural economic models. Annual Review of Economics. 5(1). 457-486.
- Keane, M.P. 2010. Structural vs. atheoretic approaches to econometrics. Journal of Econometrics. 156(1). 3–20.
- Wald , A. 1950: Statistical Decision Functions. New York, NY: John Wiley & Sons.
- Andrews, I. and J.M. Shapiro. 2021. A Model of Scientific Communication. Econometrica 89. 2117-2142.

References

- V. Chernozhukov, I. Fernández-Val, A. Galichon. 2009. Improving point and interval estimators of monotone functions by rearrangement. Biometrika 96(3). 559–575.
- Conley, T.G., C.B. Hansen, and P.E. Rossi. 2012. Plausibly exogenous. Review of Economics and Statistics 94(1). 260-272.
- Nunn, N. and L. Wantchekon. 2011. The Slave Trade and the Origins of Mistrust in Africa. American Economic Review. 101(7). 3221-3252.
- Andrews, I., M. Gentzkow, and J.M. Shapiro. 2017. Measuring the Sensitivity of Parameter Estimates to Estimation Moments. Quarterly Journal of Economics. 1553-1592

References

- Berry, S., J. Levinsohn, and A. Pakes. 1995. Automobile prices in market equilibrium. Econometrica 63(4): 841-890.
- Gourinchas, P-O and J.A . Parker. 2002. Consumption over the life cycle. Econometrica 70(1): 47-89.
- DellaVigna, S, J.A. List, and U. Malmendier. 2012. Testing for altruism and social pressure in charitable giving. Quarterly Journal of Economics 127(1): 1-56.
- Ottonello, P. and Winberry, T. 2020. Financial Heterogeneity and the Investment Channel of Monetary Policy. Econometrica 88: 2473-2502.

References

- Chen, Z., Z. Liu, J.C. Suárez Serrato, and D.Y. Xu. 2021. Notching R&D Investment with Corporate Income Tax Cuts in China. American Economic Review 111(7): 2065–2100.
- Balke, N. and T. Lamadon. 2022. Productivity Shocks, Long-Term Contracts, and Earnings Dynamics. American Economic Review 112 (7): 2139–77.
- Bernard, A.B., E. Dhyne, G. Magerman, K. Manova, and A. Moxnes. 2022. The Origins of Firm Heterogeneity: A Production Network Approach. Journal of Political Economy 130(7). 1765-1804
- Kreindler, G. 2024. Peak-Hour Road Congestion Pricing: Experimental Evidence and Equilibrium Implications. Econometrica 92. 1233-1268.

References

- Gentzkow, M. 2007. Valuing New Goods in a Model with Complementarity: Online Newspapers. American Economic Review 97(3). 713–744.
- Hendren, N. 2013. Private Information and Insurance Rejections. Econometrica 81: 1713-1762.
- Attanasio, O.P., C. Meghir, and A. Santiago. 2012. Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. The Review of Economic Studies 79(1). 37–66.
- Andrews, I., M. Gentzkow, and J.M. Shapiro. 2020. On the Informativeness of Descriptive Statistics for Structural Estimates. Econometrica 88. 2231-2258.

References

- Goldsmith-Pinkham, P., I. Sorkin, and H. Swift. 2020. Bartik Instruments: What, When, Why, and How. American Economic Review. 110(8). 2586–2624.
- Bartik, T.J. 1991. Who Benefits from State and Local Economic Development Policies? Kalamazoo, MI: W.E. Upjohn Institute.
- Goldsmith-Pinkham, P. 2024. Tracking the Credibility Revolution across Fields. Working paper.
- Adao, R., M. Kolesár, and E. Morales. 2019. Shift-Share Designs: Theory and Inference. Quarterly Journal of Economics. 134(4). 1949–2010.
- K. Borusyak, P. Hull, X. Jaravel. 2022. Quasi-Experimental Shift-Share Research Designs. The Review of Economic Studies. 89(1). 181–213.

References

- Honoré, B., T. Jørgensen, Á. de Paula. 2020. The informativeness of estimation moments. J Appl Econ. 35. 797–813.
- Jørgensen, T.H. 2023. Sensitivity to Calibrated Parameters. The Review of Economics and Statistics. 105(2). 474–481.
- Armstrong, T.B. and M. Kolesár. 2021. Sensitivity analysis using approximate moment condition models. Quantitative Economics 12. 77-108.
- Bonhomme, S. and M. Weidner. 2022. Minimizing sensitivity to model misspecification. Quantitative Economics 13. 907-954.
- Christensen, T. and B. Connault. 2023. Counterfactual Sensitivity and Robustness. Econometrica 91. 263-298.
- Staiger, D., and J.H. Stock. 1997. Instrumental Variables Regression with Weak Instruments. Econometrica 65. 557–586.

References

- Andrews, D.W.K. 1999. Estimation When a Parameter is on a Boundary. Econometrica 67. 1341–1383.
- Cont, R. 2001. Empirical properties of asset returns: stylized facts and statistical issues. Quantitative Finance 1(2). 223-236
- Hirano, K., and J.R. Porter. 2012. Impossibility Results for Nondifferentiable Functionals. Econometrica 80. 1769–1790.
- Andrews, I., and A. Mikusheva. 2016. A Geometric Approach to Nonlinear Econometric Models. Econometrica 84. 1249–1264.
- Andrews, I. and J.M. Shapiro. 2024. Communicating Scientific Uncertainty via Approximate Posteriors (Previously: Bootstrap Diagnostics for Irregular Estimands). Working paper.