A Prediction Approach to Structural Identification

Jason Weitze

Department of Economics
Stanford University

Metrics Workshop

March 10th, 2025

Motivation

Setting: A researcher has a model & wants to answer an economic or policy question

E.g., In a life-cycle savings model, how much are consumers saving as informal insurance?

Typical First Q: Do my chosen moments identify the model?

E.g., Do consumption moments identify the life-cycle savings model?

Problem: Identification analysis is Hard $\Rightarrow$ Status quo relies on heuristic arguments

> In complex dynamic models, it is well known that formal proofs of semiparametric identification are infeasible. So it is standard to rely on heuristic arguments.
>
> Eckstein, Keane & Lifshitz (2019 ECMA)

This Paper: Propose a flexible, computational alternative to traditional identification analysis

Outlining Our Approach

1. Model broader decision problem motivating identification analyses
   i. Researchers choose data to estimate quantity of interest
   ii. Characterizes a researcher's value for different types of data
      → Value of population data gives measure of identification

2. Translate decision problem into an algorithmic prediction exercise
   - Researcher's value for a given type of data = how well it predicts their quantities of interest
   - By combining simulations with machine learning we can measure the value of data

What we get:

- Algorithmic framework for directly choosing data to estimate quantity of interest
- For other use cases, we can still answer identification questions

Researcher's Decision Problem: Gourinchas and Parker (2002)

Setting:

Model: Life-cycle savings model

Objective (Loosely): What motivates savings behavior across life-cycle?

The Researcher's Decision Problem

0. Choose what data to observe e.g., consumption moments vs. savings moments?
1. Observe chosen moments (consumption moments)
2. Provide estimates on relative importance of savings mechanisms across life-cycle
3. Receive utility based on quality of estimates

"But, I already have data"

→ Identification questions conceptually precede the data

→ Later, we extend framework to incorporate existing data (closer to local identification)

Researcher's Decision Problem

Setting:

- Model: $\theta \in \Theta_0$ specifies a (parametric) distribution $\mathbb{P}_\theta(X)$

  $X_k \sim N(f_k(\theta), 1)$ for known f_k e.g., $f_1(\theta) = \theta^3 - \frac{5}{3}\theta^2 + \theta$

- Objective: Learn about an economic or policy-relevant quantity, $C(\theta)$

  $C(\theta) = \mathbf{1}\{\theta > c\}$

The Researcher's Decision Problem

0. Choose which moments, $M_n(\theta) \in \mathcal{M}(\theta)$, to observe

   $M_{n,k}(\theta) = \mathbb{E}_n[X_k(\theta)]$ for $k = 1, 2$

1. Observe chosen moments, $M_n(\theta) = m_n$
2. Provide estimates $\delta(m_n)$ of $C(\theta)$
3. Receive utility, $U(C(\theta), \delta(m_n)) = -(C(\theta) - \delta(m_n))^2$

Researcher's Decision Problem & Identification

Setting:

- Model: $\theta \in \Theta_0$ specifies a (parametric) distribution $\mathbb{P}_\theta(X)$

  $X_k \sim N(f_k(\theta), 1)$ for known f_k e.g., $f_1(\theta) = \theta^3 - \frac{5}{3}\theta^2 + \theta$

- Objective: Identify θ, $\mathbf{C(\theta) = \theta}$

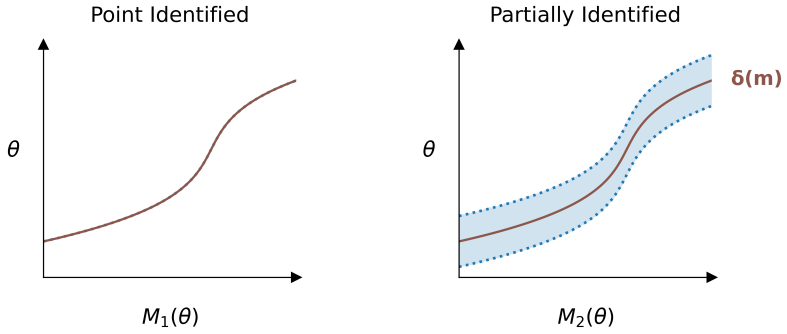
The Researcher's Decision Problem

0. Choose which **population** moments, $M(\theta) \in \mathcal{M}(\theta)$, to observe

   $M_k(\theta) = \mathbb{E}[X_k(\theta)]$ for $k = 1, 2$

1. Observe chosen **population** moments, $M(\theta) = m$
2. Provide estimates $\delta(m)$ of $\boldsymbol{\theta}$
3. Receive utility, $U(\boldsymbol{\theta}, \delta(m))$

Researcher's Decision Problem & Identification: An Illustration

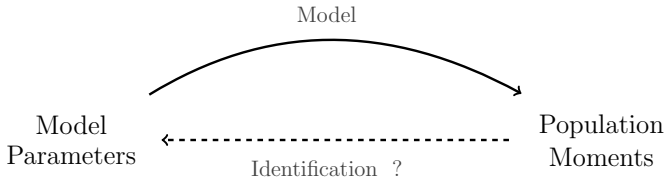


Point identification $\iff$ Perfect answers given population data

Why is Identification Hard to Study?

Model gives us moment functions, $M(\theta)$ (or at least we can approximate with simulations)

$\rightarrow$ But identification concerns it's inverse, which we often don't know



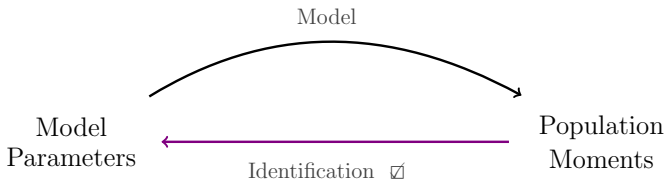
This asymmetry may motivate minimum distance estimation

$$\hat{\theta}(m_n) = \min_{\theta} d(M(\theta), m_n)$$

Why is Identification Hard to Study?

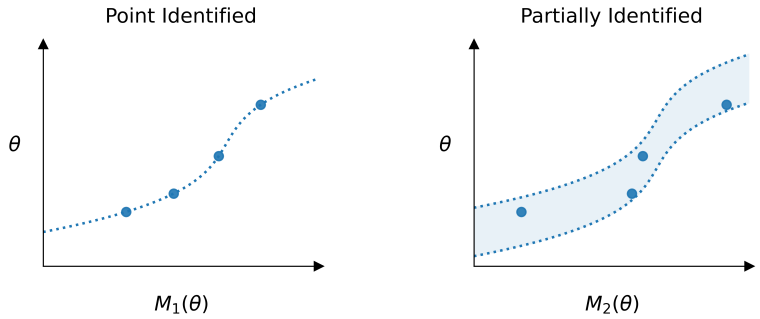
Model gives us moment functions, $M(\theta)$ (or at least we can approximate with simulations)

$\rightarrow$ But identification concerns it's inverse, which we often don't know



This Paper: Use knowledge of $M(\theta)$ to learn $\hat{\delta}(m)$

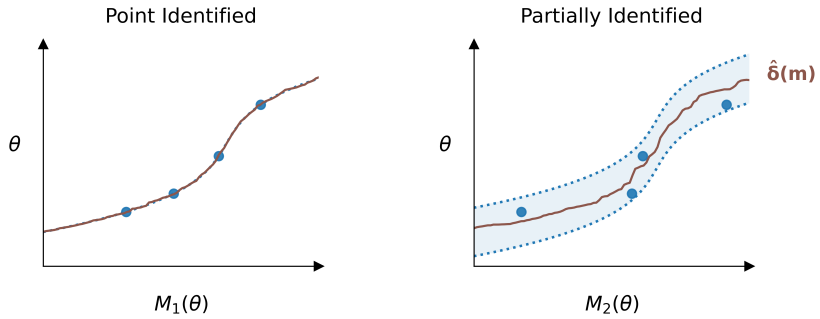
Our Approach



Leverage knowledge of $M(\theta)$

i. Pick $\theta_s \in \Theta_0$, and compute its population moments, $M(\theta_s)$ for $s \in \{1, \ldots, S\}$

Our Approach

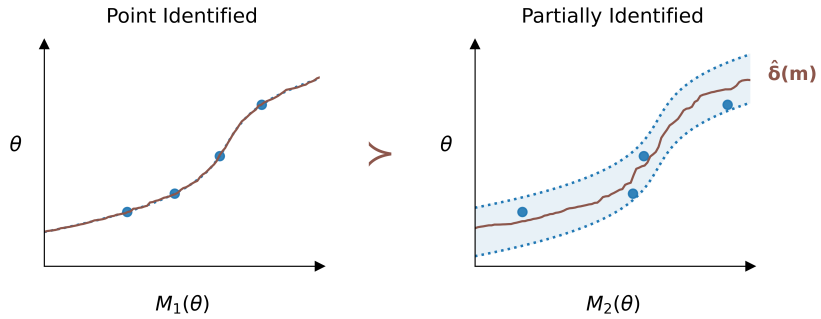


Leverage knowledge of $M(\theta)$

i. Pick $\theta_s \in \Theta_0$, and compute its population moments, $M(\theta_s)$ for $s \in \{1, \ldots, S\}$

ii. Use prediction tools to learn to predict θ given $M(\theta)$

Our Approach



Leverage knowledge of $M(\theta)$

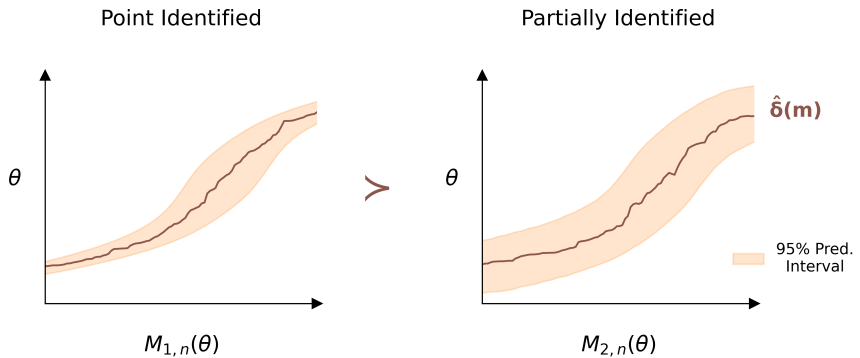
i. Pick $\theta_s \in \Theta_0$, and compute its population moments, $M(\theta_s)$ for $s \in \{1, \ldots, S\}$

ii. Use prediction tools to learn to predict θ given $M(\theta)$

iii. Assess out of sample utility to approximate the value of the data

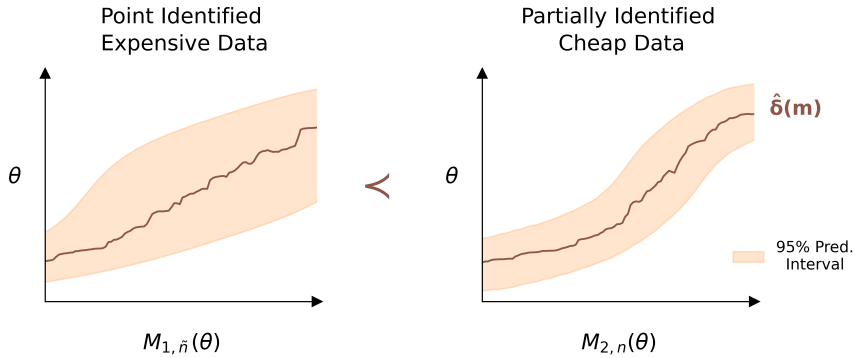
Model Identification is not Necessarily the Researcher's Goal

a. Identification concerns population, but typically observe samples, e.g., $M_n(\theta)$



Model Identification is not Necessarily the Researcher's Goal

a. Identification concerns population, but typically observe samples, e.g., $M_n(\theta)$

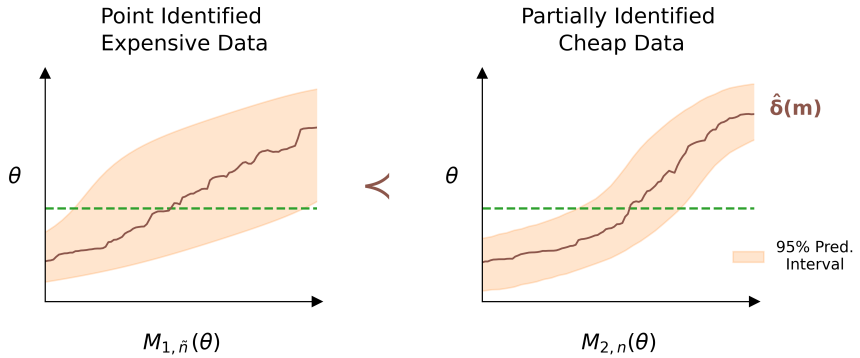


Model Identification is not Necessarily the Researcher's Goal

a. Identification concerns population, but typically observe samples, e.g., $M_n(\theta)$

b. Researchers often interested in $C(\theta)$, not θ

E.g., $C(\theta) = \mathbf{1}\{\theta > c\}$

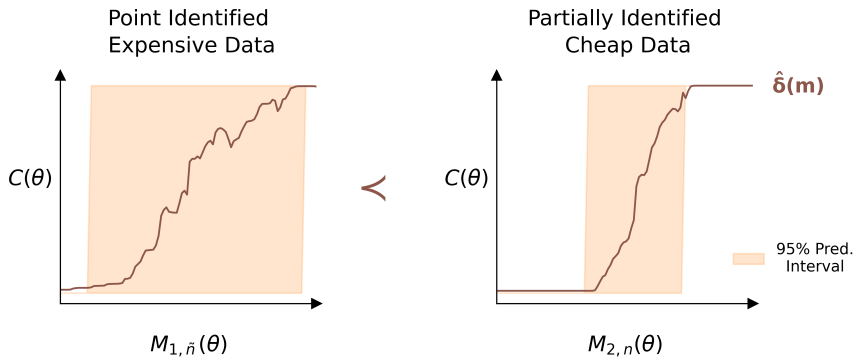


Model Identification is not Necessarily the Researcher's Goal

a. Identification concerns population, but typically observe samples, e.g., $M_n(\theta)$

b. Researchers often interested in $C(\theta)$, not θ

E.g., $C(\theta) = \mathbf{1}\{\theta > c\}$



Roadmap

Roadmap

Characterizing The Value of Data

Working Backwards:

- → Assume researchers chose $M_n(\theta) \in \mathcal{M}(\theta)$ and have prior $\pi(\theta)$
- → Then they pick δ to maximize expected utility

$$\delta^* \;=\; \arg\max_{\delta \in \Delta} \; \mathbb{E}_{\theta \sim \pi} \, U\big(\, C(\theta)\,,\, \delta(M_n(\theta))\,\big)$$

Characterizing The Value of Data

Working Backwards:

- → Assume researchers chose $M_n(\theta) \in \mathcal{M}(\theta)$ and have prior $\pi(\theta)$
- → Then they pick δ to maximize expected utility

$$\delta^* = \arg\max_{\delta \in \Delta} \text{ Average Utility } \left(\delta \, ; M_n \right)$$

How do researchers' value data?

- → Based on quality of subsequent decisions

$$V(M_n) = \text{ Average Utility } \left(\delta^* \, ; M_n \right)$$

The Researcher's Decision Problem:

$$M_n^* = \arg\max_{M_n \in \mathcal{M}_n} V(M_n)$$

Interpreting The Researcher's Value of Data

If $V(M_n)$ doesn't have interpretable units, we suggest a simple transformation

$\rightarrow$ How much does M_n improve upon trivial moments relative to oracle moments?

$M^{\text{trivial}}(\theta) = 1$ $\qquad$ $M^{\text{oracle}}(\theta) = \theta$

General Case:

$$\widetilde{V}(M_n) = \frac{V(M_n) - V(M^{\text{trivial}})}{V(M^{\text{oracle}}) - V(M^{\text{trivial}})}$$

Interpreting The Researcher's Value of Data

If $V(M_n)$ doesn't have interpretable units, we suggest a simple transformation

$\rightarrow$ How much does M_n improve upon trivial moments relative to oracle moments?

$M^{\text{trivial}}(\theta) = 1$ $\qquad$ $M^{\text{oracle}}(\theta) = \theta$

Special Case, RMSE:

$$I(M_n) = \frac{StD_\pi(\, C(\theta)\,) \;-\; RMSE_\pi\,(\, C(\theta) \,|\, M_n(\theta)\,)}{StD_\pi(\, C(\theta)\,)}$$

Interpretations:

- Observing M_n is expected to explain $I(M_n)$% of the prior StD
- $I(M) = 1 \iff$ Point Identification (w.p. 1)

Roadmap

Measuring the Value of Data: $V(M_n)$

Problem: we don't know δ, or $\mathbb{E}_\theta$

$$V(M_n) = \mathbb{E}_{\theta \sim \pi} \left[U \left(C(\theta), \delta(M_n) \right) \right]$$

Measuring the Value of Data: $V(M_n)$

Second Best: Use estimated counterpart,

$$\hat{V}_{s_0,s_1}(M_n) \;=\; \hat{\mathbb{E}}_{s_1}\left[U\left(C(\theta)\,,\,\hat{\delta}_{s_0}(M_n)\right)\right]$$

Measuring the Value of Data: $V(M_n)$

Second Best: Use estimated counterpart,

$$\hat{V}_{S_0,S_1}(M_n) \;=\; \hat{\mathbb{E}}_{S_1}\left[U\left(C(\theta)\,,\,\hat{\delta}_{S_0}(M_n)\right)\right]$$

Algorithm:

1. For $i = 1, \ldots, S = S_0 + S_1$:
   a. Draw $\theta_i \sim \pi$
   b. Compute moments $M_n(\theta_i) = m_{n,i}$ and quantity of interest $C(\theta_i) = c_i$

Measuring the Value of Data: $V(M_n)$

Second Best: Use estimated counterpart,

$$\hat{V}_{S_0,S_1}(M_n) \;=\; \hat{\mathbb{E}}_{S_1}\left[U\left(\,C(\theta)\,,\,\hat{\delta}_{S_0}(M_n)\right)\right]$$

Algorithm:

1. For $i = 1, \ldots, S = S_0 + S_1$:
   a. Draw $\theta_i \sim \pi$
   b. Compute moments $M_n(\theta_i) = m_{n,i}$ and quantity of interest $C(\theta_i) = c_i$

2. Estimate $\hat{\delta}_{S_0}$ using first S_0 obs:

$$\text{e.g.,} \qquad \hat{\delta}_{S_0} = \arg\min_{\delta \in \Delta_S} \; S_0^{-1} \sum_i \left(\, c_i \;-\; \delta(m_{n,i})\,\right)^2$$

Measuring the Value of Data: $V(M_n)$

Second Best: Use estimated counterpart,

$$\hat{V}_{S_0,S_1}(M_n) \;=\; \hat{\mathbb{E}}_{S_1}\Big[U\Big(\,C(\theta)\,,\,\hat{\delta}_{S_0}(M_n)\Big)\Big]$$

Algorithm:

1. For $i = 1, \ldots, S = S_0 + S_1$:
   a. Draw $\theta_i \sim \pi$
   b. Compute moments $M_n(\theta_i) = m_{n,i}$ and quantity of interest $C(\theta_i) = c_i$

2. Estimate $\hat{\delta}_{S_0}$ using first S_0 obs:

$$\text{e.g.,} \qquad \hat{\delta}_{S_0} = \arg\min_{\delta \in \Delta_S} \; S_0^{-1} \sum_i \big(\, c_i \;-\; \delta(m_{n,i})\,\big)^2$$

3. Compute out-of-sample utility using the other S_1 obs

Formal Properties of Algorithm

Properties of Algorithm

If $\hat{\delta}_{S_0}$ is a consistent estimate for δ^* (as $S_0, S_1 \to \infty$), then:

i. $\hat{V}_{S_0,S_1}(M_n)$ is consistent for $V(M_n)$

ii. $\hat{V}_{S_0,S_1}(M_n)$ is an unbiased lower bound for $V(M_n)$

Implications of lower bound for identification analyses:

- For fixed S, can prove point-identification, but not partial-identification
- Fortunately, researchers rarely want to prove partial identification
- Researchers arguing point-identification are incentivized to $\uparrow S$ until they persuade critics

Roadmap

An Empirical Application: Gourinchas and Parker (2002)

- Objective: Why do consumers save across the life-cycle?
- Model: Agents trade-off consumption today (C_t) and in the future

$$V_t(W_t, P_t) = \max_{C_t} \frac{C_t^{1-\rho}}{1-\rho} + \beta \mathbb{E}\left[V_{t+1}(W_{t+1}, P_{t+1})\right]$$

- Where:
  - W_t is cash on hand (i.e., liquid assets)
  - P_t is the permanent component of income
- Income grows randomly, but savings grow deterministically
- Boundary Condition: Agents consume C_{T+1} in retirement,

$$C_{T+1} = \gamma_0 P_{T+1} + \gamma_1 W_{T+1}$$

Identify & Estimate Model with Avg. Consumption Profile

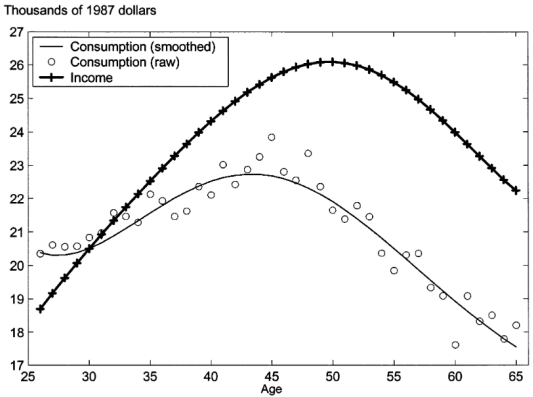


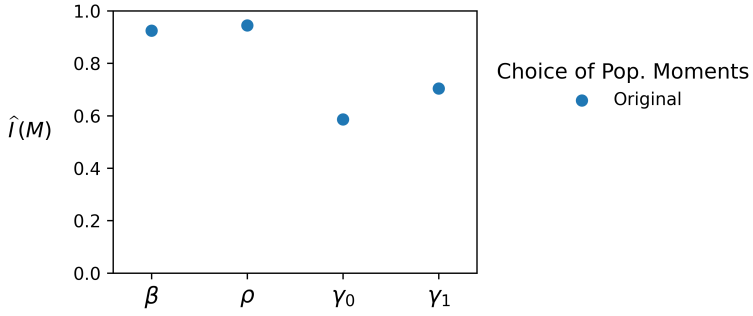
FIGURE 2.—Household consumption and income over the life cycle.

Q: How well does the consumption profile predict the parameters?

Can we Predict the Parameters?

Simulate moments from $S = 120,000$ different parameter values, θ_i

- For each $\theta_i = (\beta_i, \rho_i, \gamma_{0,i}, \gamma_{1,i})$, simulate $n = 100,000$ agents
- Original Moments = avg. cons. profile, $M_n(\theta_i) = \left\{\mathbb{E}_{n,\theta_i}[C_t] : t = 26, \ldots, 65\right\}$



Next Steps in Gourinchas and Parker

#0. Have we run enough simulations?

#1. Accept possibility of partial identification
   a. E.g., Estimate identified set

#2. Choose additional moments to make γ_0, γ_1 more predictable

#3. Are the retirement motives easier to predict?

Next Steps in Gourinchas and Parker

#0. Have we run enough simulations?

#1. Accept possibility of partial identification
   a. E.g., Estimate identified set

#2. Choose additional moments to make γ_0, γ_1 more predictable

#3. Are the retirement motives easier to predict?

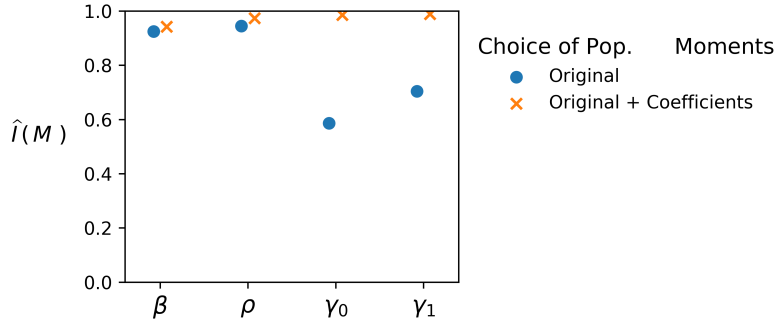
2. Can Additional Moments Improve Predictability?

Retirement consumption function: $C_t = \gamma_0 P_t + \gamma_1 W_t$

$\rightarrow$ If we observed P_t, estimate regression coefs

Idea: Use Y_t as a proxy for P_t and use estimated coefs as 'moments'

$$C_t = \beta_0 Y_t + \beta_1 W_t + \varepsilon_t$$



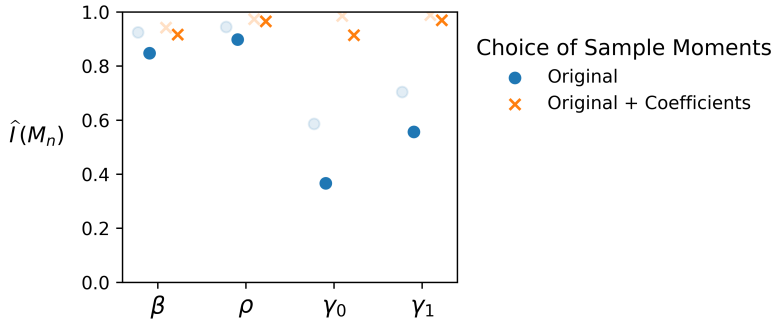
2. Can Additional Moments Improve Predictability?

Retirement consumption function: $C_t = \gamma_0 P_t + \gamma_1 W_t$

$\rightarrow$ If we observed P_t, estimate regression coefs

Idea: Use Y_t as a proxy for P_t and use estimated coefs as 'moments'

$$C_t = \beta_0 Y_t + \beta_1 W_t + \varepsilon_t$$



Next Steps in Gourinchas and Parker

#0. Have we run enough simulations?

#1. Accept possibility of partial identification

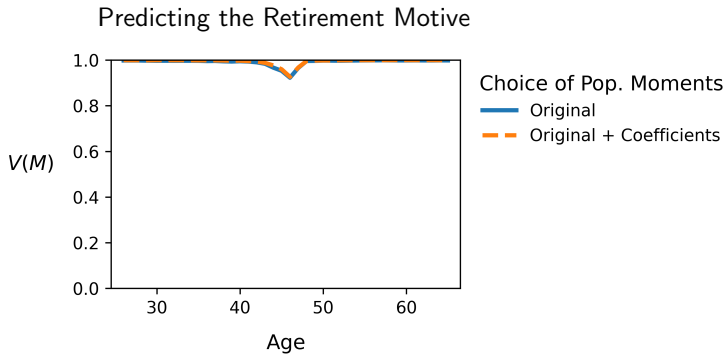
   a. E.g., Estimate identified set

#2. Choose additional moments to make γ_0, γ_1 more predictable

#3. Are the retirement motives easier to predict?

3. Is the Retirement Motive Predictable?

Authors Q: When are Retirement Savings > Precautionary Savings?



Take-Away #1: Relationship between savings motives is identified by both sets of moments

3. Is the Retirement Motive Predictable?

Authors Q: When are Retirement Savings > Precautionary Savings?

Take-Away #2: Relationship between savings motives is predictable without data (i.e., it's mechanical)

Roadmap

Taking Stock

Recap

1. We consider researchers choosing moments to maximize $V(M_n)$
2. We can measure $V(M_n)$ for a given choice of M_n: $\hat{V}_{S_0,S_1}(M_n)$
3. Choose moments with largest $\hat{V}_{S_0,S_1}(M_n)$

Follow-Up Exercises

a. If we already have data, we can incorporate it
   $\rightarrow$ Can compute local value of data (analagous to local identification)

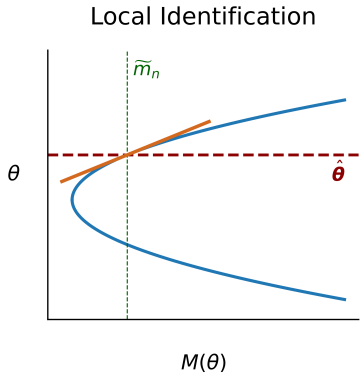
b. If $\mathcal{M}(\theta)$ is large, we can maximize $V(M_n)$ more computationally efficiently
   $\rightarrow$ Separately evaluating $V(M_n)$ for each $M_n \in \mathcal{M}(\theta)$ can be prohibitively expensive

Incorporating Existing Data

Researcher has observed $\widetilde{M}_n(\theta_0) = \widetilde{m}_n$

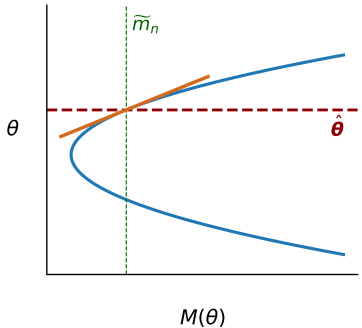
- New Decision Problem: Choose which moments to observe next
- Characterizes local value of data: $V(M_n|\widetilde{m}_n)$ = "Incremental value of $M_n(\theta)$ given $\widetilde{m}_n$"

Local Value of Data and Local Identification

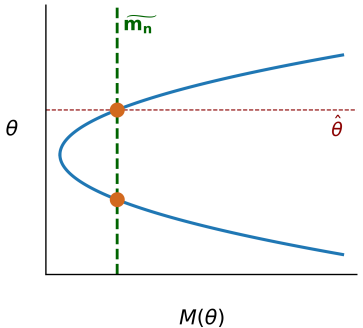


Local Value of Data and Local Identification

Local to Parameters



Local to Moments



Incorporating Existing Data

Researcher has observed $\widetilde{M}_n(\theta_0) = \widetilde{m}_n$

- New Decision Problem: Choose which moments to observe next
- Characterizes local value of data: $V(M_n|\widetilde{m}_n)$ = "Incremental value of $M_n(\theta)$ given $\widetilde{m}_n$"

How? Replace prior with posterior: $\pi(\theta|\widetilde{m}_n)$

→ We provide algorithm that efficiently leverages simulated data to estimate posterior

Typical Q: Which moments of my dataset, D, should I use?

- Requires estimating $\pi(\theta|D)$, which is preferred to one based on moments
  → But often difficult to estimate
- Alt: Iteratively condition on moments, and ask, 'which additional moments?'
  → Can select moments s.t. sequence of posteriors converge to $\pi(\theta|D)$

Dealing with a Large Choice Space, $\mathcal{M}(\theta)$

1. Parameterize moment space by growing class of prediction functions
   - Consider $\mathcal{M}(\theta) = \{\mathbb{E}_n[\phi(X(\theta))] \,:\, \phi \in \Phi\}$
   - Let Φ_S be a class of prediction functions that approximates Φ as $S \to \infty$
   - Operationally, we learn $\phi : X(\theta) \mapsto C(\theta)$

     $\Rightarrow$ Choose moments, $M_n(\theta) = \mathbb{E}_n\left[\hat{\phi}_{S_0}(X(\theta))\right]$

2. If choice space is large, but discrete, select moments using nonlinear lasso

$$\delta, \beta = \max_{\delta,\beta} \text{Average Utility}\left(\delta\,; M_n\,\text{diag}(\beta)\right) - \lambda^{cv}||\beta||_1$$

$\rightarrow$ Implicitly selects moments with $\beta_k \neq 0$, i.e., $M_n = \{M_{n,k} \,:\, \beta_k \neq 0\}$

Roadmap

Conclusion

(Feedback? Email: jweitze@stanford.edu, Thanks!)

Conceptual Framework:

1. Embed identification in a broader decision problem:
   "What data should we collect to inform our policy question?"
2. Translate this decision problem into an algorithmic prediction exercise
   Intuition: Value of data $\approx$ how well it predicts $C(\theta)$

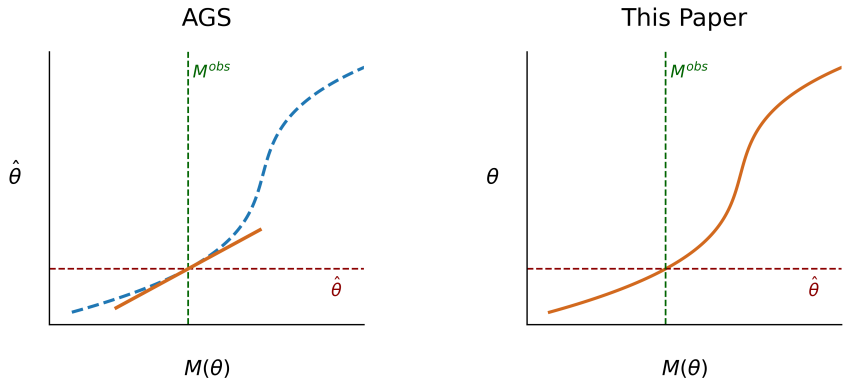
Algorithmic Framework:

3. Combine simulations & machine learning to measure predictive ability of data, and thus, its value
4. Generalize core algorithm to:
   a. Incorporate existing data
   b. Choose what data to collect more efficiently

Empirical Application:

5. Life-cycle savings model in Gourinchas and Parker (2002)

Relation to AGS (Andrews, Gentzkow & Shapiro, 2017)



Key Differences:

- AGS provide local sensitivity measure
  - → AGS: How do (small) changes in M map into $\hat{\theta}$?
  - → This Paper: How does M map into θ?