Automatic Debiased Machine Learning of Structural Parameters with General Conditional Moments

Facundo Argañaraz

Universidad Carlos III de Madrid

Aarhus University

January 21, 2025

Outline

1 Introduction and Motivation

2 A Motivating Example

3 Literature Review

4 (Recap of) Debiased Machine Learning

5 Construction of OR-IVs

6 Estimation and Inference

7 Some general results (short discussion)

8 Empirical Application

9 Final Remarks

Introduction and Motivation

This paper

- General models defined by a finite number of conditional moment restrictions (CMRs).
- CMRs are allowed to depend on non-parametric components η_0.
- CMRs are allowed to depend on finite-dimensional parameters θ_0 non-linearly.
- I develop a method for conducting inference on θ_0 when η_0 is estimated by machine learning.

Why Machine Learning (ML)?

- ML tools have shown great potential for handling the estimation of unknown high-dimensional components η_0 without imposing parametric specifications.
    - Particularly, for prediction.
- Many economic models entail η_0. This typically needs to be estimated before proceeding with the estimation/inference of θ_0, the parameter of interest.
- Then, ML tools have huge potential within our economic models for CAUSAL INFERENCE

Why Machine Learning (ML)?

- ML tools have shown great potential for handling the estimation of unknown high-dimensional components η_0 without imposing parametric specifications.
  - Particularly, for prediction.
- Many economic models entail η_0. This typically needs to be estimated before proceeding with the estimation/inference of θ_0, the parameter of interest.
- Then, ML tools have huge potential within our economic models for CAUSAL INFERENCE... but (unfortunately) ML introduces statistical challenges.

A Motivating Example: Production Functions

Example: Production Functions I

- A panel of n firms across T periods is observed, where i and t index firms and periods, respectively.
- Let Y_{it} be the (log of) output of firm i at time t, and X_{it} be a vector of (log of) inputs, e.g., capital and labor.
- Output is

$$Y_{it} = F\left(X_{it}, \theta_{0p}\right) + \omega_{it} + \epsilon_{it}, \tag{1}$$

  - F is assumed to be known up to θ_{0p}, e.g., a Cobb-Douglas specification in logs.
  - ω_{it} is firm i's productivity shock in period t, which is allowed to be correlated with inputs.
  - ϵ_{it} is noise in output (independent of everything).

Example: Production Functions II

- Proxy variable approach: Olley and Pakes (1996); see also Levinsohn and Petrin (2003) and Wooldridge (2009).
- Suppose that we observe I_t—e.g., investment or intermediate inputs—such that

$$\omega_{it} = \omega_t\left(I_{it}, X_{it}\right),$$

  where $\omega_t(\cdot)$ is unknown.
- Equation (1) becomes

$$Y_{it} = F\left(X_{it}, \theta_{0p}\right) + \omega_t\left(I_{it}, X_{it}\right) + \epsilon_{it}.$$

Example: Production Functions III

- Let $\eta_{0t}(I_{it}, X_{it}) = F(X_{it}, \theta_{0p}) + \omega_t(I_{it}, X_{it})$. Then,

$$\mathbb{E}\left[Y_{it} - \eta_{0t}(I_{it}, X_{it}) \middle| I_{it}, X_{it} \right] = 0.$$

- Assume that ω_{it} follows a First-Order Markov's process in the sense that (Ackerberg et al., 2014)

$$\mathbb{E}\left[\omega_{it} \middle| \omega_{i,t-1} \right] = \theta_{0\omega} \omega_{i,t-1}.$$

- Let Ω_{it} be the firm i's information set at t. It is not difficult to show that

$$\mathbb{E}\left[Y_{it} - F(X_{it}, \theta_{0p}) - \theta_{0\omega}\left(\eta_{0,t-1}(Z_{i,t-1}) - F(X_{i,t-1}, \theta_{0p})\right) \middle| \Omega_{i,t-1} \right] = 0.$$

Production Functions IV

- Suppose that $T = 3$.
- The model can be defined by the following CMRs:

$$
\begin{aligned}
\mathbb{E}\left[Y_1 - \eta_{01}\left(I_1, X_1\right) \middle| I_1, X_1 \right] &= 0, \\
\mathbb{E}\left[Y_2 - F\left(X_2, \theta_{0p}\right) - \theta_{0\omega}\left(\eta_{01}\left(I_1, X_1\right) - F\left(X_1, \theta_{0p}\right)\right) \middle| \Omega_1 \right] &= 0, \\
\mathbb{E}\left[Y_2 - \eta_{02}\left(I_2, X_2\right) \middle| I_2, X_2 \right] &= 0, \\
\mathbb{E}\left[Y_3 - F\left(X_3, \theta_{0p}\right) - \theta_{0\omega}\left(\eta_{02}\left(I_2, X_2\right) - F\left(X_2, \theta_{0p}\right)\right) \middle| \Omega_2 \right] &= 0.
\end{aligned}
$$

- We want to estimate $\theta_0 = \left(\theta_{0p}', \theta_{0\omega}\right)'$ when $\eta_0 = (\eta_{01}, \eta_{02})$ is unknown too.
- Then, what would a researcher do?

Production Functions IV

- Suppose (for now) that we follow a *naive* approach.

1. **First Stage:** Estimate η_0 using $\mathbb{E}\left[\left. Y_t - \eta_{0t}\left(I_t, X_t\right)\right| I_t, X_t\right] = 0$ by:
   - OLS of Y_t on I_t, X_t, I_t^2, X_t^2, I_tX_t,..., like in Ackerberg et al. (2015) (ACF).
   - Instead, I propose to use more flexible modern tools, e.g., Boosting, Random Forest, Lasso, Neural Networks, and the like.
2. **Second Stage:** Plug fitted values of the previous stage and estimate θ_0 by GMM using:

$$\mathbb{E}\left[\left(Y_2 - F\left(X_2, \theta_{0p}\right) - \theta_{0\omega}\left(\eta_{01}\left(I_1, X_1\right) - F\left(X_1, \theta_{0p}\right)\right)\right) \otimes f\left(\Omega_1\right)\right] = 0,$$

$$\mathbb{E}\left[\left(Y_3 - F\left(X_3, \theta_{0p}\right) - \theta_{0\omega}\left(\eta_{02}\left(I_2, X_2\right) - F\left(X_2, \theta_{0p}\right)\right)\right) \otimes f\left(\Omega_2\right)\right] = 0.$$

What happens if we reproduce this exercise 1,000 times and check the distribution of $\frac{\sqrt{n}\left(\hat{\theta}-\theta_0\right)}{s.e.(\hat{\theta})}$? Is this asymptotically $N(0,1)$?

Figure: Comparison of naive procedure and my procedure

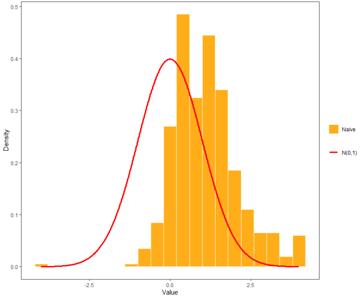
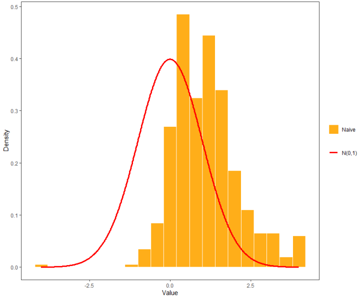
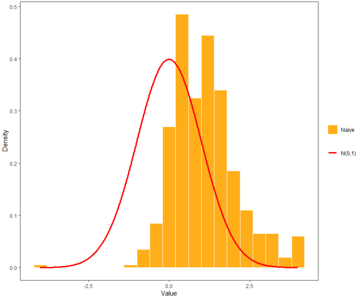


Figure: Comparison of naive procedure and my procedure

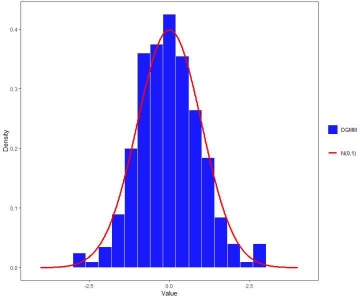


CAUTION: This simple plug-in procedure (naive procedure) will not yield valid inference on θ_0!

Figure: Comparison of Naive and DGMM Estimators



(a) Naive



(b) My procedure

Literature Review: What is new here?

What is new here? I

- This paper contributes to the body of research on CMRs, e.g., Chamberlain (1992), Brown and Newey (1998), Ai and Chen (2003), Ai and Chen (2012), Ackerberg et al. (2014), and Chen and Santos (2018).
    - While the works dealing with estimation typically propose efficient estimators that are LR when first steps are estimated by sieves (Chen, 2007), I allow for the utilization of a wide array of machine learning tools.
- This paper is also connected to the recently developed literature on debiased moments, e.g., Chernozhukov, Escanciano, Ichimura, Newey, and Robins (2022a).
    - Chernozhukov et al. (2016) characterize Orthogonal IVs (OR-IVs) in our context, but do not provide a feasible algorithm.
    - Chernozhukov et al. (2018) focus on a setting with one conditioning information set, and again do not provide a feasible algorithm.

What is new here? II

- Argañaraz and Escanciano (2024) characterize OR-IVs in our context for general functionals, and again do not provide a feasible algorithm.

- This paper contributes to the literature on the automatic construction of debiased moments; e.g., Chernozhukov et al. (2022b) and Bakhitov (2022).
    - These methods ARE NOT applicable to the setting of this paper.
        1. I do not *only* focus on situations where $\theta_0 = \mathbb{E}\left[m\left(W, \eta_0\right)\right]$.
        2. General models with semiparametric CMRs.
        3. My procedure obtains OR-IVs, instead of Riesz representers (a non-trivial difference).

What is new here? III

- To the best of my knowledge, no previous paper has used off-the-shelf machine learning techniques for estimating production functions while maintaining the ability to perform standard inference.
  - Cha et al. (2023) develop an adjusted DML estimator of Chernozhukov et al. (2018), but focus on only one strategy to estimate nuisance components.

Conducting Inference on Structural Parameters when Machine Learning is Used

LR Moments I

- Locally Robust (LR)/Orthogonal/Debiased Moment (Chernozhukov et al., 2022a, 2018; Neyman, 1959): $\psi : \mathcal{W} \times \Theta \times \mathbf{B} \times L^2(Z) \mapsto \mathbb{R}$

$$\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0, \kappa_0\right)\right] = 0,$$

where κ_0 is a "special" vector of functions of the conditioning variables (exogenous variables) in the model that has finite second moment. These are **Orthogonal IVs (OR-IVs)**.

LR Moments I

- Locally Robust (LR)/Orthogonal/Debiased Moment (Chernozhukov et al., 2022a, 2018; Neyman, 1959): $\psi : \mathcal{W} \times \Theta \times \mathbf{B} \times L^2(Z) \mapsto \mathbb{R}$

$$\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0, \kappa_0\right)\right] = 0,$$

where κ_0 is a "special" vector of functions of the conditioning variables (exogenous variables) in the model that has finite second moment. These are **Orthogonal IVs (OR-IVs)**.

- LR moments lead to valid inference when machine learning is used to estimate nuisance parameters because

$$(i)\ \frac{d}{d\tau}\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0 + \tau b, \kappa_0\right)\right] = 0, \quad \text{for all } b \in \mathbf{B},$$

$$(ii)\ \mathbb{E}\left[\psi\left(W, \theta_0, \eta_0, \kappa\right)\right] = 0, \quad \text{for all } \kappa \in L^2(Z).$$

LR Moments II

- We can obtain a debiased moment by means of (Argañaraz and Escanciano, 2024)

LR Moments II

- We can obtain a debiased moment by means of (Argañaraz and Escanciano, 2024)

$$
\begin{aligned}
\psi\left(W,\theta_0,\eta_0,\kappa_0\right) = &\left(Y_1-\eta_{01}\left(I_1,X_1\right)\right)\kappa_{01}\left(Z_1\right)\\
&+\left(Y_2-F\left(X_2,\theta_{0p}\right)-\theta_{0\omega}\left(\eta_{01}\left(Z_1\right)-F\left(X_1,\theta_{0p}\right)\right)\right)\kappa_{02}\left(Z_1\right)\\
&+\left(Y_2-\eta_{02}\left(Z_2\right)\right)\kappa_{03}\left(Z_2\right)\\
&+\left(Y_3-F\left(X_3,\theta_{0p}\right)-\theta_{0\omega}\left(\eta_{02}\left(Z_2\right)-F\left(X_2,\theta_{0p}\right)\right)\right)\kappa_{04}\left(Z_2\right),
\end{aligned}
$$

where $Z_1=(I_1,X_1)$, $Z_2=(I_2,X_2)$.

LR Moments II

- We can obtain a debiased moment by means of (Argañaraz and Escanciano, 2024)

$$
\begin{aligned}
\psi\left(W,\theta_0,\eta_0,\kappa_0\right) = {} & \left(Y_1 - \eta_{01}\left(I_1, X_1\right)\right)\kappa_{01}\left(Z_1\right) \\
& + \left(Y_2 - F\left(X_2,\theta_{0p}\right) - \theta_{0\omega}\left(\eta_{01}\left(Z_1\right) - F\left(X_1,\theta_{0p}\right)\right)\right)\kappa_{02}\left(Z_1\right) \\
& + \left(Y_2 - \eta_{02}\left(Z_2\right)\right)\kappa_{03}\left(Z_2\right) \\
& + \left(Y_3 - F\left(X_3,\theta_{0p}\right) - \theta_{0\omega}\left(\eta_{02}\left(Z_2\right) - F\left(X_2,\theta_{0p}\right)\right)\right)\kappa_{04}\left(Z_2\right),
\end{aligned}
$$

  where $Z_1 = (I_1, X_1)$, $Z_2 = (I_2, X_2)$.

- $\kappa_0 = \left(\kappa_{01}, \kappa_{02}, \kappa_{03}, \kappa_{04}\right) \in L^2(Z)$ is such that

$$
(i)\ \frac{\partial}{\partial\eta}\mathbb{E}\left[\psi\left(W,\theta_0,\eta,\kappa_0\right)\right] = 0.
$$

LR Moments II

- We can obtain a debiased moment by means of (Argañaraz and Escanciano, 2024)

$$\begin{aligned}\psi\left(W,\theta_0,\eta_0,\kappa_0\right) = &\left(Y_1-\eta_{01}\left(I_1,X_1\right)\right)\kappa_{01}\left(Z_1\right)\\ &+\left(Y_2-F\left(X_2,\theta_{0p}\right)-\theta_{0\omega}\left(\eta_{01}\left(Z_1\right)-F\left(X_1,\theta_{0p}\right)\right)\right)\kappa_{02}\left(Z_1\right)\\ &+\left(Y_2-\eta_{02}\left(Z_2\right)\right)\kappa_{03}\left(Z_2\right)\\ &+\left(Y_3-F\left(X_3,\theta_{0p}\right)-\theta_{0\omega}\left(\eta_{02}\left(Z_2\right)-F\left(X_2,\theta_{0p}\right)\right)\right)\kappa_{04}\left(Z_2\right),\end{aligned}$$

  where $Z_1=(I_1,X_1)$, $Z_2=(I_2,X_2)$.
- $\kappa_0=\left(\kappa_{01},\kappa_{02},\kappa_{03},\kappa_{04}\right)\in L^2(Z)$ is such that

$$(i)\ \frac{\partial}{\partial\eta}\mathbb{E}\left[\psi\left(W,\theta_0,\eta,\kappa_0\right)\right]=0.$$

- By the law of iterated expectations, *(ii)* holds too.

LR Moments II

- We can obtain a debiased moment by means of (Argañaraz and Escanciano, 2024)

$$\begin{aligned}\psi(W,\theta_0,\eta_0,\kappa_0) = {} & (Y_1 - \eta_{01}(I_1,X_1))\,\kappa_{01}(Z_1) \\ & + (Y_2 - F(X_2,\theta_{0p}) - \theta_{0\omega}(\eta_{01}(Z_1) - F(X_1,\theta_{0p})))\,\kappa_{02}(Z_1) \\ & + (Y_2 - \eta_{02}(Z_2))\,\kappa_{03}(Z_2) \\ & + (Y_3 - F(X_3,\theta_{0p}) - \theta_{0\omega}(\eta_{02}(Z_2) - F(X_2,\theta_{0p})))\,\kappa_{04}(Z_2),\end{aligned}$$

  where $Z_1 = (I_1, X_1)$, $Z_2 = (I_2, X_2)$.

- $\kappa_0 = \left(\kappa_{01}, \kappa_{02}, \kappa_{03}, \kappa_{04}\right) \in L^2(Z)$ is such that

$$(i)\ \frac{\partial}{\partial\eta}\mathbb{E}\left[\psi(W,\theta_0,\eta,\kappa_0)\right] = 0.$$

- By the law of iterated expectations, *(ii)* holds too.
- **Thus, if we want a LR moment, we need to obtain $\boldsymbol{\kappa_0}$.**

LR Moments III

- Obtaining OR-IVs is not trivial.
  1. They are the solution to potentially challenging-to-solve functional equations.
  2. They might depend on unknown population objects such as θ_0 and η_0.
- **I propose a feasible approach to estimate OR-IVs that leads to a debiased moment.**
- The researcher does not have to explicitly solve functional equations and unknown objects are properly dealt with $\rightarrow$ "automatic" or data-driven approach.

How can we get κ_0?

(Main Contribution)

How can we get κ_0? I

- Compute derivatives of each CMR:

$$\left[S^{(1)}_{\theta_0,\eta_0} b\right](Z_1) = -b_1(Z_1), \quad \left[S^{(2)}_{\theta_0,\eta_0} b\right](Z_1) = -\theta_{0\omega} b_1(Z_1),$$
$$\left[S^{(3)}_{\theta_0,\eta_0} b\right](Z_2) = -b_2(Z_2), \quad \left[S^{(4)}_{\theta_0,\eta_0} b\right](Z_2) = -\theta_{0\omega} b_2(Z_2).$$

- Notice that each of the above is a linear operator.
- Collect these derivatives in the linear operator:

$$S_{\theta_0,\eta_0} b = \left(S^{(1)}_{\theta_0,\eta_0} b, S^{(2)}_{\theta_0,\eta_0} b, S^{(3)}_{\theta_0,\eta_0} b, S^{(4)}_{\theta_0,\eta_0} b\right).$$

- For a valid κ_0 we need

$$\frac{d}{d\tau} \mathbb{E}\left[\psi\left(W, \theta_0, \eta + \tau b, \kappa_0\right)\right] = \sum_{j=1}^{4} \mathbb{E}\left[\left[S^{(j)}_{\theta_0,\eta_0} b\right](Z_j) \kappa_{0j}(Z_j)\right] = 0.$$

How can we get κ_0? II

- Technically, κ_0 is orthogonal to $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$.
  - $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$ is a subspace of $L^2(Z)$, a Hilbert space with inner product $\langle f, h\rangle = \sum_{j=1}^{4} \mathbb{E}\left[f\left(Z_j\right) h_j\left(Z_j\right)\right]$
- If $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)} \neq L^2(Z)$, pick some function $f \in L^2\left(Z\right)$,e.g., $f(Z) = Z$ and compute the residual

$$\kappa_0 = f - \Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}} f.$$

where $\Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}}$ denotes the orthogonal projection operator onto $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$ (or "fitted values").

How can we get κ_0? III

- Notice that $\Pi_{\overline{\mathcal{R}\left(S_{\theta,\eta_0}\right)}} f \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$.
- The range of $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is dense in $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$, where $S^*_{\theta_0,\eta_0}$ is the adjoint operator of S_{θ_0,η_0}.
- It satisfies, for all $g \in \mathbf{B}$,

$$\left\langle S_{\theta_0,\eta_0} b, g \right\rangle_{L^2(Z)} = \left\langle b, S^*_{\theta_0,\eta_0} g \right\rangle_{\mathbf{B}} .$$

- Then, $\Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}} f \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}\right)}$.

How can we get κ_0? IV

- $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is an operator that maps elements in $L^2(Z)$ into $L^2(Z)$, spaces of exogenous variables in the data.
- $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ can be characterized by exploiting the CMRs given by the model in a general way.
- In general, we are interested in solving

$$\inf_{g \in L^2(Z)} \sum_{j=1}^{4} \mathbb{E}\left[\left(f_j(Z_j) - S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0} g\right)^2\right]. \tag{2}$$

- But...
  - $S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is unknown $\rightarrow$ Estimate it.
  - A search over an infinite-dimensional space $\rightarrow$ Regularize the solution.

Estimation of κ_0

Estimation of κ_0 I

- We will look for a solution to (2) in a set $\mathcal{G}_n$ defined as

$$\mathcal{G}_n := \left\{ g \in L^2(Z) : g_j\left(Z_j\right) = \gamma_j\left(Z_j\right)^{'} \beta_j, \;\; j = 1, \cdots, 4, \;\; ||\beta||_\infty < c \right\}.$$

- In practice, we solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{4} \mathbb{E}_n \left[\left(f_j(Z_{ji}) - \hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}} \gamma(Z_i)^{'} \beta \right)^2 \right] + 2\lambda_n \, ||\beta||_1 \,, \qquad (3)$$

where $\gamma(Z)^{'}\beta \equiv \left(\gamma_1(Z_1)^{'}\beta_1, \cdots, \gamma_J(Z_J)^{'}\beta_J \right)^{'}$.

- The solution to (3) is denoted by $\hat{\beta}$.
- More compactly, in the sample, $\hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}} \gamma(Z_i)$ is a vector of r dimension, so let me $\hat{M}_j(Z_{ji}) \equiv \hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}} \gamma(Z_i)$.

Estimation of κ_0 II

- In practice, we solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{4} \mathbb{E}_n \left[\left(f_j(Z_{ji}) - \hat{M}_j(Z_{ji})'\beta \right)^2 \right] + 2\lambda_n \left|\left|\beta\right|\right|_1 .$$

- Observe that $\hat{\beta}$ solves a regularized version of

$$\sum_{j=1}^{4} \mathbb{E}_n \left[\hat{M}_j(Z) \left(f_j(Z) - \hat{M}_j(Z)'\hat{\beta} \right) \right] = 0,$$

  which aims to assure property *(i)* of LR moment, i.e., zero derivative with respect to η_0.
- This is nothing but a constrained OLS problem or a Lasso program. We know how to solve Lasso-type programs!

Estimation of κ_0 III

- Ultimately, I propose to estimate each component in κ_0, by means of

$$\begin{aligned}\hat{\kappa}_j\left(Z_{ji}\right) &= f_j\left(Z_{ji}\right) - \hat{f}_j^*\left(Z_{ji}\right) \\ &= f_j\left(Z_{ji}\right) - \hat{M}_j(Z_{ji})'\hat{\beta}, \quad j = 1, \cdots, 4.\end{aligned}$$

Algorithm I

The algorithm that I propose is:

1. Start with a known (vector-) function of the conditioning variables, say $f(Z) = (f_1(Z_1), \cdots, f_4(Z_4))'$.
2. Construct suitable vectors $\hat{M}_j(Z_j)$ that estimates the derivatives of the sample moment of $\psi(W, \hat{\theta}, \hat{\eta}, \hat{\kappa})$ (with respect to η_0).
3. Get an estimator of an OR-IV by means of

$$\hat{\kappa}_j(Z_j) = f_j(Z_j) - \hat{M}(Z_j)'\hat{\beta},$$

where $\hat{\beta}$ "solves"

$$\sum_{j=1}^{4} \mathbb{E}_n \left[\hat{M}_j(Z_j) \left(f_j(Z_j) - \hat{M}_j(Z_j)'\hat{\beta} \right) \right] = 0.$$

and $\hat{\beta}$ can be seen as the estimated coefficient of a Lasso program.

Algorithm II

- $\hat{M}(Z)$ estimates the derivatives of the sample moment of $\psi(W, \hat{\theta}, \hat{\eta}, \hat{\kappa})$ (with respect to η).
- In our example,

$$
\begin{aligned}
\hat{M}_1(Z_{1i}) &= \left(\gamma_1(Z_{1i}) + \hat{\theta}_\omega \gamma_1(Z_{1i}), \cdots, \gamma_r(Z_{1i}) + \hat{\theta}_\omega \gamma_r(Z_{1i})\right), \\
\hat{M}_2(Z_{1i}) &= \hat{\theta}_\omega \hat{M}_1(Z_{1i}), \\
\hat{M}_3(Z_{2i}) &= \left(\gamma_1(Z_{2i}) + \hat{\theta}_\omega \gamma_1(Z_{2i}), \cdots, \gamma_r(Z_{2i}) + \hat{\theta}_\omega \gamma_r(Z_{2i})\right), \\
\hat{M}_4(Z_{2i}) &= \hat{\theta}_\omega \hat{M}_3(Z_{1i}).
\end{aligned}
$$

- Notice: Regressors depend on $\hat{\theta}_\omega$. It might be the case that they also depend on other nuisance parameters that we can estimate, e.g., some conditional expectations.

Estimator of the Parameter of Interest

- Let $\hat{\eta}_\ell$ and $\hat{\boldsymbol{\kappa}}_\ell$ be estimators obtained by *cross-fitting* (Chernozhukov et al., 2018, and references therein).
  - The sample is divided in subsamples $\ell = 1, \cdots, L$.
  - $\hat{\eta}_\ell$ and $\hat{\boldsymbol{\kappa}}_\ell$ are obtained from observations not in subsample ℓ.
- Let
$$\hat{\psi}(\theta) = \frac{1}{n} \sum_{\ell=1}^{L} \sum_{i \in I_\ell} \psi\left(W_i, \theta, \hat{\eta}_\ell, \hat{\boldsymbol{\kappa}}_\ell\right).$$
- The proposed estimator $\hat{\theta}$ is simply defined as the solution to the GMM program:
$$\hat{\theta} = \arg\min_{\theta \in \Theta} \hat{\psi}(\theta)' \hat{\Lambda} \hat{\psi}(\theta), \tag{4}$$
where $\hat{\Lambda}$ is a positive semi-definite weighting matrix.

That's it?

That's it?
(Well... not really)

More in the paper I

- In the paper, I actually deal with a general setting: ▸ More details

$$\mathbb{E}\left[m_j\left(Y,\theta_0,\eta_0\right)\middle|\, Z_j\right]=0, \quad \text{a.s.}, \quad j=1,\cdots,J.$$

- η_0 is allowed to depend on variables not in Z_j, as in cases with endogeneity.
- The algorithm to estimate OR-IVs works in this general situation.

  ▸ More details
- The Debiased GMM (DGMM) estimator $\hat{\theta}$ is introduced in a two-step framework.

More in the paper II

- I show:

Claim

Given some ▸ *sufficient conditions* *in the paper,*

$$||\hat{\kappa} - \kappa_0||_{L^2(Z)} = O_p\left(\mu_n\right), \quad \textit{for a specific } \mu_n \xrightarrow[n \to \infty]{} 0.$$

where $||f||_{L^2(Z)} = \sqrt{\sum_{j=1}^{J} ||f_j||_2^2}$.

More in the paper III

- I show:

Claim

Given some sufficient conditions *in the paper,*

$$\sqrt{n}\left(\hat{\theta}-\theta_0\right) \xrightarrow{d} N\left(0,V\right), \quad V=\left(\Upsilon'\Lambda\Upsilon\right)^{-1}\Upsilon'\Lambda\Psi\Lambda\Upsilon\left(\Upsilon'\Lambda\Upsilon\right)^{-1}.$$

$\hat{V} \xrightarrow{p} V$, *where*

$$\hat{\Psi}=\frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\hat{\psi}_{i\ell}\hat{\psi}_{i\ell}', \quad \hat{\psi}_{i\ell}\equiv\psi\left(W_i,\tilde{\theta}_\ell,\hat{\eta}_\ell,\hat{\boldsymbol{\kappa}}_\ell\right),$$

$$\hat{\Upsilon}=\frac{\partial}{\partial\theta}\hat{\psi}(\hat{\theta}).$$

Empirical Application: How can my approach help practitioners?

Data

- Data from the Instituto Nacional de Estadistica de Chile.
- Census of Chilean manufacturing plants with at least ten employees and I focus on the period 1985-1987.
- A very popular database in the literature of production functions (e.g., Ackerberg et al., 2015; Alvarez and López, 2005; Gandhi et al., 2020; Levinsohn and Petrin, 2003; Pavcnik, 2002).
- I have information on revenues, capital, labor, intermediate inputs,...
- Sample consists of firms from the five largest three-digit ISIC manufacturing industries in Chile: food products, textiles, apparel, wood products, and fabricated metal products. I pool all sectors.
- In total, after data cleaning, 619 establishments.

Empirical Application I

- Examine a production function in two inputs: L_{it} and K_{it}, labor and capital, respectively.
- Cobb-Douglas specification (in logs):

$$Y_{it} = \theta_{01} + \theta_{0l} L_{it} + \theta_{0k} K_{it} + \omega_{it} + \epsilon_{it}. \tag{5}$$

- As a proxy variable we now use intermediate inputs E_{it}, following Levinsohn and Petrin (2003).
- Note that E_{it} does not appear in (5), meaning that Y_{it} is "value-added" rather than output.

Empirical Application II

- Our starting vectors of functions are, for $2 \leq t \leq T$,

$$f_{1t}(Z) = (K_{t-1}, K_{t-1})',$$

$$f_{2t}(Z) = (K_{t-1}, K_t)',$$

$$f_{3t}(Z) = (L_{t-1}, L_{t-1})',$$

$$f_{4t}(Z) = \left(K_{t-1}^2, K_t^2\right)',$$

$$f_{5t}(Z) = \left(K_{t-1}^4, K_t^4\right)'.$$

Empirical Application III

Table: Production Function Estimation in Empirical Application

Coeff.
$\hat{\theta}_l$
$\hat{\theta}_k$
$\hat{\theta}_l + \hat{\theta}_k$
$\hat{\theta}_k / \hat{\theta}_l$

Empirical Application III

Table: Production Function Estimation in Empirical Application

Coeff.	OLS
$\hat{\theta}_l$	0.84
	(0.03)
$\hat{\theta}_k$	0.37
	(0.01)
$\hat{\theta}_l + \hat{\theta}_k$	1.21
	(0.02)
$\hat{\theta}_k / \hat{\theta}_l$	0.43
	(0.03)

Empirical Application III

Table: Production Function Estimation in Empirical Application

Coeff.	OLS	ACF ($d = 2$)
$\hat{\theta}_l$	0.84	0.34
	(0.03)	(0.26)
$\hat{\theta}_k$	0.37	0.39
	(0.01)	(0.19)
$\hat{\theta}_l + \hat{\theta}_k$	1.21	0.73
	(0.02)	(0.23)
$\hat{\theta}_k / \hat{\theta}_l$	0.43	1.14
	(0.03)	(1.27)

Empirical Application III

Table: Production Function Estimation in Empirical Application

Coeff.	OLS	ACF ($d=2$)	ACF ($d=5$)
$\hat{\theta}_l$	0.84	0.34	0.14
	(0.03)	(0.26)	(0.36)
$\hat{\theta}_k$	0.37	0.39	0.52
	(0.01)	(0.19)	(0.23)
$\hat{\theta}_l + \hat{\theta}_k$	1.21	0.73	0.66
	(0.02)	(0.23)	(0.29)
$\hat{\theta}_k / \hat{\theta}_l$	0.43	1.14	3.73
	(0.03)	(1.27)	(10.80)

Empirical Application III

Table: Production Function Estimation in Empirical Application

Coeff.	OLS	ACF ($d = 2$)	ACF ($d = 5$)	DGMM
$\hat{\theta}_l$	0.84	0.34	0.14	0.42
	(0.03)	(0.26)	(0.36)	(0.14)
$\hat{\theta}_k$	0.37	0.39	0.52	0.18
	(0.01)	(0.19)	(0.23)	(0.06)
$\hat{\theta}_l + \hat{\theta}_k$	1.21	0.73	0.66	0.59
	(0.02)	(0.23)	(0.29)	(0.10)
$\hat{\theta}_k / \hat{\theta}_l$	0.43	1.14	3.73	0.42
	(0.03)	(1.27)	(10.80)	(0.01)

Empirical Application IV

Table: Estimation Results on Productivity

Coeff.
75/25
90/10
95/05

Empirical Application IV

Table: Estimation Results on Productivity

Coeff.	OLS
75/25	3.03
	(0.11)
90/10	8.63
	(0.09)
95/05	17.07
	(0.19)

Empirical Application IV

Table: Estimation Results on Productivity

Coeff.	OLS	ACF $(d = 2)$
75/25	3.03	3.84
	(0.11)	(0.28)
90/10	8.63	11.40
	(0.09)	(0.10)
95/05	17.07	24.47
	(0.19)	(1.70)

Empirical Application IV

Table: Estimation Results on Productivity

Coeff.	OLS	ACF ($d = 2$)	ACF ($d = 5$)
75/25	3.03	3.84	3.82
	(0.11)	(0.28)	(0.13)
90/10	8.63	11.40	11.81
	(0.09)	(0.10)	(0.23)
95/05	17.07	24.47	26.89
	(0.19)	(1.70)	(0.59)

Empirical Application IV

Table: Estimation Results on Productivity

Coeff.	OLS	ACF ($d = 2$)	ACF ($d = 5$)	DGMM
75/25	3.03	3.84	3.82	4.76
	(0.11)	(0.28)	(0.13)	(0.25)
90/10	8.63	11.40	11.81	16.93
	(0.09)	(0.10)	(0.23)	(0.25)
95/05	17.07	24.47	26.89	41.27
	(0.19)	(1.70)	(0.59)	(2.19)

Final Remarks

Final Remarks

- I have introduced an approach to conduct standard inference on structural parameters when machine learning is used, within a general class of semiparametric models defined by CMRs.
- The method will hopefully pave the way for the use of machine learning techniques in contexts where the construction of LR has remained unexplored, such as highly non-linear and complex settings with CMRs, prominent in economics.
- Exciting new avenues of research for the future:
  1. Identification.
  2. Selection of more efficient OR-IVs (debiased moments).
  3. Application to other relevant empirical settings.

Thank you!

argafacu.github.io

Appendix

Monte Carlo I

- Production Function Estimation.
- Firms are followed during three periods, i.e., $T = 3$.
- Cobb-Douglas production function in logs:

$$Y_{it} = \theta_{01} + \theta_{0k} K_{it} + \omega_{it} + \epsilon_{it},$$

- where $\theta_{01} = 0$ and $\theta_{0k} = 1$.
- The law of motion of capital (in levels) is given by

$$k_{it} = (1 - \delta)\, k_{i,t-1} + \mu_{it} i_{i,t-1},$$

- where $1 - \delta = 0.9$, μ_{it} is a lognormal standard shock to the capital accumulation process, and i_{it} is the firm's investment decision.

Monte Carlo II

- The log of investment is determined by

$$I_{it} = \gamma_0 + \gamma_1 K_{it} + \gamma_2 \omega_{it} + \exp\left(-0.5 K_{it} + 0.5 \omega_{it}\right), \tag{6}$$

  where $\gamma_0 = 0$, $\gamma_1 = -0.7$, and $\gamma_2 = 5$. Note that (6) implies that $\omega_{it} = \omega_t\left(I_{it}, K_{it}\right)$.
- Productivity is assumed to follow a normal AR(1) process with $\theta_{0\omega} = 0.7$.
- We will then perturb Equation (6) by introducing a shock.

Monte Carlo III

- We automatically construct four debiased moments, and thus we have to provide four vectors of functions $f(Z)$:

$$f_1(Z) = (K_1, K_1, K_2, K_2)', \quad f_2(Z) = (I_1, I_1, I_2, I_2)',$$
$$f_3(Z) = (K_1, K_1, I_2, I_2)', \quad f_4(Z) = (K_1, I_1, I_2, I_2)'.$$

- The first stage η_{0t} is estimated by Boosting using $L = 5$ folds.
- I have computed a naive plug-in estimator (PI) as a benchmark.

More details

Table: Monte Carlo Results

n	PI Bias	DGMM Bias	PI SE	DGMM SE	PI RMSE	DGMM RMSE	PI 95Cvg	DGMM 95Cvg
				DGP 1: No shock to investment				
250	0.02	-0.03	0.12	0.14	0.12	0.14	0.88	0.92
500	0.04	-0.03	0.10	0.11	0.11	0.12	0.86	0.92
750	0.06	-0.02	0.09	0.10	0.11	0.11	0.81	0.93
$1,000$	0.06	-0.03	0.07	0.09	0.10	0.09	0.79	0.93
				DGP 2: Shock to investment (sd=0.5)				
250	0.04	-0.00	0.12	0.15	0.13	0.15	0.89	0.92
500	0.06	-0.00	0.10	0.12	0.12	0.12	0.84	0.93
750	0.08	0.00	0.09	0.11	0.12	0.11	0.77	0.95
$1,000$	0.09	-0.00	0.08	0.10	0.12	0.10	0.70	0.95
				DGP 3: Shock to investment (sd=0.7)				
250	0.06	0.02	0.13	0.16	0.14	0.16	0.88	0.93
500	0.09	0.03	0.10	0.13	0.13	0.14	0.79	0.95
750	0.10	0.03	0.08	0.13	0.13	0.13	0.69	0.97
$1,000$	0.11	0.03	0.08	0.12	0.14	0.12	0.60	0.96

General Setting I

- The data $W_i = (Y_i, X_i, Z_i)$, $i = 1, \cdots, n$ is idd with the same distribution as $W = (Y, X, Z)$, where Y has support $\mathcal{Y} \in \mathbb{R}^{d_Y}$.
- $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ denotes a finite-dimensional parameter vector.
- $\eta \in \Xi \subseteq \mathbf{B}$ be a vector of real-valued measurable functions of X.
- $\mathbf{B} \subseteq \bigotimes^{d_\eta} L^2(X)$ is some Hilbert space with inner product $\langle \cdot, \cdot \rangle_{\mathbf{B}}$.
- To be specific, $\eta = \left(\eta_1, \cdots, \eta_{d_\eta}\right)$ with $\eta_s \equiv \eta_s(X)$.

General Setting II

- Residual functions $m_j : \mathcal{Y} \times \Theta \times \mathbf{B} \mapsto \mathbb{R}$ such that:

$$\mathbb{E}\left[m_j\left(Y, \theta_0, \eta_0\right) \middle| Z_j \right] = 0, \quad \text{a.s.}, \quad j = 1, 2, \cdots, J, \tag{7}$$

- Z denotes the union of distinct random elements of the conditioning variables variables $(Z_1, \cdots, Z_J)$.
- Hereafter, I assume that there exists a unique $(\theta_0, \eta_0) \in \Theta \times \mathbf{B}$ such that (7) holds.
- m_j might depend on θ_0 arbitrarily and the entire function η_0, not only on its evaluation at a particular realization of X.
- I say the vector Z gathers the exogenous variables in the model, which might include X or not.
- Let $\kappa = (\kappa_1, \cdots, \kappa_J)$, where $\kappa_j \equiv \kappa_j\left(Z_j\right)$, and $\kappa_j \in L^2(Z_j)$, $1 \leq j \leq J$. Hence, $\kappa \in L^2(Z)$, where $L^2(Z) = \bigotimes_{j=1}^{J} L^2(Z_j)$. I define

$$h_j\left(Z_j, \theta, \eta\right) := \mathbb{E}\left[m_j\left(Y, \theta, \eta\right) \middle| Z_j \right].$$

General Setting III

- We maintain a key assumption throughout:

Assumption

For all j, $h_j\left(Z_j, \theta_0, \cdot\right) : \mathbf{B} \mapsto L^2(Z_j)$ is Fréchet differentiable in a neighborhood of η_0, where the derivative $S^{(j)}_{\theta_0,\eta_0} b$ is given by

$$\begin{aligned} \left[\nabla h_j\left(Z_j, \theta_0, \eta_0\right)\right](b) &\equiv \frac{d}{d\tau} h_j\left(Z_j, \theta_0, \eta_0 + \tau b\right) \\ &= \left[S^{(j)}_{\theta_0,\eta_0} b\right]\left(Z_j\right), \end{aligned} \tag{8}$$

for all $b \in \mathbf{B}$.

- Remark that (8) defines a linear operator $S^{(j)}_{\theta_0,\eta_0} : \mathbf{B} \mapsto L^2(Z_j)$; see Luenberger (1997) and Carrasco and Florens (2000) for a theory on linear operators.

General Setting IV

- In addition, let us define

$$S_{\theta_0,\eta_0} b = \left(S^{(1)}_{\theta_0,\eta_0} b, \cdots, S^{(J)}_{\theta_0,\eta_0} b \right).$$

- Then, $S_{\theta_0,\eta_0} : \mathbf{B} \mapsto L^2(Z)$ is also a linear operator. We equip $L^2(Z)$ with the inner product $\langle f_1(Z), f_2(Z) \rangle_{L^2(Z)} = \sum_{j=1}^{J} \mathbb{E}\left[f_{1j}(Z_j) f_{2j}(Z_j) \right]$, where $f_s = (f_{s1}, \cdots, f_{sJ})$, $s = 1, 2$.
- Therefore, $L^2(Z)$ is a Hilbert space. The range of that operator can be defined as follows

$$\mathcal{R}\left(S_{\theta_0,\eta_0}\right) = \left\{ f \in L^2(Z) : f = S_{\theta_0,\eta_0} b \text{ for some } b \in \mathbf{B} \right\}.$$

- A key object for us is $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}^{\perp}$, i.e., the orthogonal complement of the closure of the range of S_{θ_0,η_0} in $L^2(Z)$, which can be defined as

$$\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}^{\perp} = \left\{ f \in L^2(Z) : \sum_{j=1}^{J} \mathbb{E}\left[f_j(Z_j)\, h_j(Z_j) \right] = 0, \quad \text{for all} \quad h = (h_1, \cdots, h_J) \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)} \right\}.$$

General Setting V

- Argañaraz and Escanciano (2024) show that a debiased moment in model (7) can be constructed as follows:

$$\psi\left(W,\theta_0,\eta_0\right)=\sum_{j=1}^{J} m_j\left(Y,\theta_0,\eta_0\right)\kappa_{0j}\left(Z_j\right),\quad \kappa_0\in\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}^{\perp}. \quad (9)$$

- Pick some function $f\in L^2\left(Z\right)$, e.g., $f(Z)=Z$. Then, compute the residual

$$\kappa_0=f-\Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}}f.$$

- $\Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}}$ denotes the orthogonal projection operator onto $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$ (or "fitted values") and is defined as follows

$$\Pi_{\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}}f:=\underset{\tilde{f}\in\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}}{\arg\min}\sum_{j=1}^{J}\mathbb{E}\left[\left(f_j(Z_j)-\tilde{f}_j(Z_j)\right)^2\right].$$

General Setting VI

- Notice that $\Pi_{\overline{\mathcal{R}(S_{\theta,\eta_0})}} f \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$.
- The range of $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is dense in $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$, where $S^*_{\theta_0,\eta_0}$ is the adjoint operator of S_{θ_0,η_0}.
- It satisfies, for all $g \in \mathbf{B}$,

$$\left\langle S_{\theta_0,\eta_0} b, g \right\rangle_{L^2(Z)} = \left\langle b, S^*_{\theta_0,\eta_0} g \right\rangle_{\mathbf{B}}.$$

- Then, $\Pi_{\overline{\mathcal{R}(S_{\theta_0,\eta_0})}} f \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}\right)}$.

General Setting VII

- $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is an operator that maps elements in $L^2(Z)$ into $L^2(Z)$, spaces of exogenous variables in the data.
- $S_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ can be characterized by exploiting the CMRs given by the model in a general way.
- In general, we are interested in solving

$$\inf_{g \in L^2(Z)} \sum_{j=1}^{J} \mathbb{E}\left[\left(f_j(Z_j) - S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0} g\right)^2\right]. \tag{10}$$

- But...
  - $S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is unknown $\rightarrow$ Estimate it.
  - A search over an infinite-dimensional space $\rightarrow$ Regularize the solution.

General Setting VIII

- We will look for a solution to (10) in a set $\mathcal{G}_n$ defined as

$$\mathcal{G}_n := \left\{ g \in L^2(Z) : g_j\left(Z_j\right) = \gamma_j\left(Z_j\right)' \beta_j, \ \ ||\beta||_\infty < c \right\}.$$

- In practice, we solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{J} \mathbb{E}_n \left[\left(f_j(Z_{ji}) - \hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S}^*{}_{\hat{\theta},\hat{\eta}} \gamma(Z_i)' \beta \right)^2 \right] + 2\lambda_n \, ||\beta||_1 \,, \qquad (11)$$

where $\gamma(Z)'\beta \equiv \left(\gamma(Z_1)'\beta_1, \cdots, \gamma(Z_J)'\beta_J \right)'$.

- The solution to (11) is denoted by $\hat{\beta}$.

Back

Algorithm to estimate OR-IVs I

- I now want to propose a specific algorithm, which can be directly followed in a wide range of settings that share a common structure in the operator S_{θ_0,η_0}. We assume

Assumption

For all j,

i) *There exists a known (up to θ_0 and η_0) function ν_j such that*

$$\left[S^{(j)}_{\theta_0,\eta_0} b\right](Z_j) = \mathbb{E}\left[\nu_j\left(Y,\theta_0,\eta_0,b\right)\middle| Z_j\right], \quad \textit{for all} \quad b \in \mathbf{B};$$

ii) $\nu_j\left(Y,\theta_0,\eta_0,b\right) = b\left(X\right)' \tilde{\nu}_j\left(Y,\theta_0,\eta_0\right)$, *for some $d_\eta-$vector of known (up to θ_0 and η_0) functions $\tilde{\nu}_j$;*

Algorithm to estimate OR-IVs II

- Leveraging Assumption 2, we can obtain an expression for $S^*_{\theta_0,\eta_0}$ that can apply to various contexts:

Proposition

Suppose Assumptions 1 and 2 hold. In addition, let $\langle b_1, b_2\rangle_{\mathbf{B}} = \mathbb{E}\left[b_1(X)' b_2(X)\right]$. *Then, the adjoint* $S^*_{\theta_0,\eta_0} : L^2(Z) \mapsto \mathbf{B}$ *exists, is linear, continuous, and given by*

$$\left[S^*_{\theta_0,\eta_0} g\right](X) = \sum_{j=1}^{J} \mathbb{E}\left[\tilde{\nu}_j\left(Y, \theta, \eta_0\right) g_j\left(Z_j\right) \middle| X\right].$$

Algorithm to estimate OR-IVs III

- Split I_ℓ^c into three mutually exclusive pieces such that $I_\ell^c = A_\ell \cup B_\ell \cup C_\ell$. Possibly, this partition depends on j, but I omit this dependence for simplicity.
- For any estimator, a subscript will indicate the part of I_ℓ^c that has been used to compute it. For example, $\hat{\theta}_{A_\ell}$ means that θ_0 has been estimated using observations in A_ℓ and so on.
- Let $\boldsymbol{f_{j\ell}}$ be a $n_\ell-$dimensional vector containing each $f_j\left(Z_{ji}\right)$, $i \notin I_\ell$.
- Let $\boldsymbol{\hat{M}_{j\ell}}$ be a $n_\ell \times r$ design matrix such that its (i,l)-entry is given by

$$\left[\boldsymbol{\hat{M}_{j\ell}}\right]_{il} = \hat{\mathbb{E}}_{C_\ell}\left[\left(\hat{\mathbb{E}}_{B_\ell}\left[\tilde{\nu}_{j'}\left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right)\gamma_{j'k}\left(Z_{ji}\right)\middle| X_i\right]\right)' \tilde{\nu}_j\left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right)\middle| Z_{ji}\right], \quad j',$$

$$j = 1, \cdots, J, \quad k = 1, \cdots, r. \tag{12}$$

Algorithm to estimate OR-IVs IV

- Then, we have

$$\hat{\beta}_\ell = \underset{\beta \in \mathbb{R}^r}{\arg\min} \sum_{j=1}^{J} \frac{1}{n - n_\ell} \left(\boldsymbol{f_{j\ell}} - \boldsymbol{\hat{M}_{j\ell}} \beta \right)' \left(\boldsymbol{f_{j\ell}} - \boldsymbol{\hat{M}_{j\ell}} \beta \right) + 2\lambda_n \, ||\beta||_1 \, ,$$

  where $\hat{\eta}_{A_\ell}$, $\hat{\eta}_{B_\ell}$, $\hat{\mathbb{E}}_{B_\ell} [\cdot | \, X]$, and $\hat{\mathbb{E}}_{C_\ell} [\cdot | \, Z_j]$ are non-parametric estimators, possibly based on some machine learning tool, and $\hat{\theta}_{A_\ell}$ and $\hat{\theta}_{B_\ell}$ are possibly non-LR estimators of θ_0.[1]

- Recall that $\tilde{\nu}_j$ is a known function, given estimators of θ_0 and η_0. Thus, these conditional expectations can be evaluated.
- Notice that the ℓ_1-penalization terms allows for $r > n$.
- The resulting residuals from such problem is an estimator of an OR-IV.

Algorithm to estimate OR-IVs V

- **Step 0:** Choose a real-valued function $f \in L^2(Z)$. Choose a basis for each $\gamma_j(Z_j)$, e.g., exponential, Fourier, splines, or power. In addition, specify a low-dimensional dictionary, say $\gamma^{low}(Z)$, which is a sub-vector of $\gamma(Z)$.
- **Step 1:** For each $\ell = 1, \cdots L$, compute (possible) non-LR estimators $\hat{\theta}_{A_\ell}$ and $\hat{\theta}_{B_\ell}$. Moreover, using some machine learning algorithm, compute $\hat{\eta}_{A_\ell}$, $\hat{\eta}_{B_\ell}$, $\hat{\mathbb{E}}_{B_\ell}[\cdot | X]$, and $\hat{\mathbb{E}}_{C_\ell}[\cdot | Z_j]$.
- **Step 2:** Construct design matrix $\hat{\boldsymbol{M}}_{\boldsymbol{j\ell}}$ such that its $(i, l)-$entry is (12).

Algorithm to estimate OR-IVs VI

- **Step 3:** Initialize $\hat{\beta}_\ell$ using $\gamma^{low}(Z)$ such that

$$\left[\hat{\boldsymbol{M}}_{j\ell}^{low}\right]_{il} = \hat{\mathbb{E}}_{C_\ell}\left[\left(\hat{\mathbb{E}}_{B_\ell}\left[\tilde{\nu}_{j'}\left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{jA_\ell}\right)\gamma_{j'k}^{low}\left(Z_{j'i}\right)\middle| X_i\right]\right)' \tilde{\nu}_j\left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{jB_\ell}\right)\middle| Z_{ji}\right],$$

$$\hat{\beta}_\ell = \begin{pmatrix} \left(\sum_{j=1}^J \hat{\boldsymbol{M}}_{j\ell}^{low\prime} \hat{\boldsymbol{M}}_{j\ell}^{low}\right)^{-1}\left(\sum_{j=1}^J \hat{\boldsymbol{M}}_{j\ell}^{low\prime} \boldsymbol{f}_{j\ell}\right) \\ 0 \end{pmatrix}.$$

- **Step 4:** (While $\hat{\beta}_\ell$ has not converged)
  (a) Update normalization

$$\hat{D}_{j'k\ell} = \left[\frac{1}{n - n_\ell}\sum_{i \notin I_\ell}\left\{\sum_{j=1}^J \hat{\mathbb{E}}_{C_\ell}\left[\left(\hat{\mathbb{E}}_{B_\ell}\left[\tilde{\nu}_{j'}\left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right)\gamma_{j'k}\left(Z_{j'i}\right)\middle| X_i\right]\right)' \tilde{\nu}_j\left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right)\middle| Z_{ji}\right]\hat{\epsilon}_{ji\ell}\right\}^2\right]^{1/2},$$

$$\hat{\epsilon}_{ji\ell} = f_j\left(Z_{ji}\right) - \hat{M}_{j\ell}\left(Z_{ji}\right)'\hat{\beta}_\ell.$$

Algorithm to estimate OR-IVs VII

(b) Update $\hat{\beta}_\ell$, where

$$\hat{\beta}_\ell = \arg\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{J} \frac{1}{n - n_\ell} \left(\boldsymbol{f}_{j\ell} - \hat{\boldsymbol{M}}_{j\ell}\beta \right)' \left(\boldsymbol{f}_{j\ell} - \hat{\boldsymbol{M}}_{j\ell}\beta \right) + 2\lambda_n \sum_{j'=1}^{J} \sum_{k=1}^{r_{j'}} \left| \hat{D}_{j'k\ell}\beta_{j'k} \right|,$$

and

$$\lambda_n = \frac{c_1}{(n - n_\ell)^{1/4}} \Phi^{-1} \left(1 - \frac{c_2}{2r} \right),$$

where $\Phi(.)$ is the standard normal cdf.

- **Step 5:** Given the optimal $\hat{\beta}_\ell$, compute $\hat{\kappa}_{j\ell}$ as:

$$\hat{\kappa}_{j\ell}(Z_{ji}) = f_j(Z_{ji}) - \hat{M}_{j\ell}(Z_{ji})' \hat{\beta}_\ell.$$

▸ Back

[1] I.e., estimators based on non-LR moments.

Coordinate Descent Approach I

- Step 4 of the iterative algorithm above requires to solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{J} \frac{1}{n - n_\ell} \left(\boldsymbol{f_{j\ell}} - \hat{\boldsymbol{M}}_{j\ell}\beta \right)' \left(\boldsymbol{f_{j\ell}} - \hat{\boldsymbol{M}}_{j\ell}\beta \right) + 2\lambda_n \left|\left| \hat{D}_\ell \beta \right|\right|_1, \quad (13)$$

- where $\hat{D}_\ell$ is a diagonal matrix with elements $\hat{D}_{jk\ell} \equiv \hat{D}_{l\ell}$ along the main diagonal, with $l = 1, \cdots, r$.
- Hence, the first r_1 entries correspond to the regressors with $\gamma_1(Z_1)$, the next r_2 entries are the regressors with $\gamma_2(Z_2)$, and so on.
- To solve (13), I use an extension of the coordinate descent approach for Lasso (Fu, 1998; Friedman et al., 2007, 2010) to our particular objective function.

Coordinate Descent Approach II

- To be precise, I implement a coordinate-wise descent algorithm with a soft-thresholding update.
- Let v_l denote the l^{th} element of a generic vector v and let e_l be a $r \times 1$ unit vector with 1 in the l^{th} coordinate and zeros elsewhere.
- This algorithm can be implemented as follows: For $l = 1 : r$, do
  1. **Step 1:** Compute loadings (which do not depend on β_k):

$$A_l = \frac{1}{n - n_\ell} \sum_{j=1}^{J} e_l' \hat{\boldsymbol{M}}_j' \left(\boldsymbol{f_j} - \hat{\boldsymbol{M}}_j \beta + \hat{\boldsymbol{M}}_j e_l \beta_l \right)$$

$$B_l = \frac{1}{n - n_\ell} \sum_{j=1}^{J} e_l' \hat{\boldsymbol{M}}_j' \hat{\boldsymbol{M}}_j e_l.$$

Coordinate Descent Approach III

2 **Step 2:** Update coordinate β_l:

$$\beta_l = \begin{cases} \frac{A_l + \hat{D}_l \lambda_n}{B_l} & \text{if} \quad A_l < -\hat{D}_l \lambda_n \\ 0 & \text{if} \quad A_l \in \left[-\hat{D}_l \lambda_n, \hat{D}_l \lambda_n\right] \\ \frac{A_l - \hat{D}_l \lambda_n}{B_l} & \text{if} \quad A_l > \hat{D}_l \lambda_n. \end{cases}$$

-end-

Back

Asymptotic results of OR-IVs I

- Let $\boldsymbol{M}_j$ be the population analog of matrix $\hat{\boldsymbol{M}}_{\boldsymbol{j\ell}}$.
- Let $\hat{M}_{j\ell}(Z_{ji})$ be a $r-$dimensional vector containing the $i-$ row of $\hat{\boldsymbol{M}}_{\boldsymbol{j\ell}}$.
- A similar definition applies to $M_j(Z_{ji})$.
- I define

$$\hat{F}_{j\ell} := \frac{1}{n-n_\ell} \sum_{i \notin I_\ell} f_j(Z_{ji}) \hat{M}_{j\ell}(Z_{ji}), \qquad F_j := \mathbb{E}\left[f_j(Z_j) M_j(Z_j)\right],$$

$$\hat{G}_{j\ell} := \frac{1}{n-n_\ell} \sum_{i \notin I_\ell} \hat{M}_{j\ell}(Z_{ji}) \hat{M}_{j\ell}(Z_{ji})', \qquad G_j := \mathbb{E}\left[M_j(Z_j) M_j(Z_j)'\right].$$

Asymptotic results of OR-IVs II

- Then, $\hat{\beta}_\ell$ can equivalently be written as

$$\hat{\beta}_\ell = \arg\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{J} \left(-2\hat{F}_{j\ell}'\beta - \beta'\hat{G}_{j\ell}\beta \right) + 2\lambda_n \|\beta\|_1 . \tag{14}$$

Assumption

There are constants $c_1, \cdots, c_J$ such that

$$\max_{1 \le k \le r} |M_{jk}(Z_j)| \le c_j, \quad a.s., \quad j = 1, \cdots, J.$$

Asymptotic results of OR-IVs III

Assumption

For all j, ℓ, k, there exists a constant C, possibly depending on j and k, such that, with probability approaching one,

$$\left|\left|\hat{M}_{j\ell k} - M_{jk}\right|\right|_2 \leq C\varepsilon_{jn}^M, \tag{15}$$

where $\varepsilon_{jn}^M \to 0$.

- Assumptions 3 and 4 imply that for any $c > 0$,

$$\left|\left|\hat{G}_{j\ell} - G_j\right|\right|_\infty = O_p\left(\varepsilon_{jn}^G\right), \quad \text{where} \quad \varepsilon_{jn}^G = \max\left\{\bar{\varepsilon}_n, n^c\varepsilon_{jn}^M\right\}, \; \bar{\varepsilon}_n = \sqrt{\frac{\log(r)}{n}}.$$

Asymptotic results of OR-IVs IV

- In our discussion to follow, let ε_n be a non-negative sequence that converges to zero no faster than ε_{jn}^G, for any j. We impose a sparse approximate condition on the orthogonal projection f^*, a key assumption for us.

Assumption

There exist $C > 1$ and $\bar{\beta} \in \mathbb{R}^r$ with s non-zero elements such that

$$\sum_{j=1}^{J} \mathbb{E}\left[\left\{f_j^*(Z_j) - M_j(Z_j)'\bar{\beta}\right\}^2\right] \leq Cs\varepsilon_n^2,$$

where $s \leq C\varepsilon_n^{-2/(2\xi+1)}$, $\xi > 0$.

Asymptotic results of OR-IVs V

Assumption

The largest eigenvalue of $\sum_{j=1}^{J} G_j$ *is uniformly bounded in* n *and there are* $C, c > 0$ *such that, for all* $\bar{s} \approx C\varepsilon_n^{-2}$,

$$\phi(\bar{s}) = \inf \left\{ \frac{\delta' \sum_j^J G_j \delta}{\left|\left|\delta_{S_\beta}\right|\right|^2}, \quad \delta \in \mathbb{R}^r \setminus \{0\}, \left|\left|\delta_{S_\beta^c}\right|\right|_1 \leq 3 \left|\left|\delta_{S_\beta}\right|\right|_1, \quad \#S_\beta \leq \bar{s} \right\} > c.$$

Assumption

For all j, *there exists a constant* C, *possibly depending on* j, *such that, with probability approaching one,* $\left|\left|\hat{F}_{j\ell} - F_j\right|\right|_\infty \leq C\varepsilon_{jn}^F$, *where* $\varepsilon_{jn}^F \to 0$.

Asymptotic results of OR-IVs VI

Assumption

Let $\varepsilon_n = \max_j \left\{ \varepsilon_{jn}^G, \varepsilon_{jn}^F \right\}$, $\varepsilon_n = o(\lambda_n)$, $\lambda_n = o(n^c \varepsilon_n)$ *for all* $c > 0$, *and there exist* $C_1 > 0$ *and* $c_2 > 1$ *such that* $r \leq C_1 n^{c_2}$.

Asymptotic results of OR-IVs VII

Theorem

Let Assumptions 3-8 hold. Then, for any $c > 0$,

$$||\hat{\kappa} - \kappa_0||_{L^2(Z)} = O_p\left(n^c \varepsilon_n^{2\xi/(2\xi+1)}\right).$$

Back

Asymptotic results of OR-IVs VIII

- The rate for the OR-IVs that we obtained relies crucially on Assumptions 4 and 7. We next provide a set of sufficient conditions for these. Additionally, they will conveniently establish a direct relationship between convergence rates of the estimators of all the unknown objects in the estimated regressors $\hat{M}_{j\ell k}$ and the rates in Assumptions 4 and 7. Let us define

$$
\begin{aligned}
R_j\left(W,\theta,\eta\right) &:= \tilde{\nu}_j\left(Y,\theta,\eta\right)\gamma_{jk}\left(Z_j\right), \\
\alpha_{0j}\left(X,\theta,\eta\right) &:= \mathbb{E}\left[\left. R_j\left(W,\theta,\eta\right)\right| X\right], \quad \alpha_{0j}\left(\cdot,\theta_0,\eta_0\right) \in \Gamma\left(X\right) \subseteq L^2(X), \\
\sigma_{0j}\left(Z_j,\alpha_{j'}\left(X,\bar{\theta},\bar{\eta}\right),\theta,\eta\right) &:= \mathbb{E}\left[\left.\alpha_{j'}\left(\bar{\theta},\bar{\eta},X\right)'\tilde{\nu}_j\left(Y,\theta,\eta\right)\right| Z_j\right], \\
&\quad \sigma_{0j}\left(\cdot,\alpha_{j'}\left(X,\theta_0,\eta_0\right),\theta_0,\eta_0\right) \in \Gamma\left(Z_j\right) \subseteq L^2(Z_j).
\end{aligned}
$$

Asymptotic results of OR-IVs IX

Assumption

i) *With probability approaching one,* $\hat{\theta}_{A_\ell}$, $\hat{\theta}_{B_\ell} \in \Theta$, $\hat{\eta}_{A_\ell}$, $\hat{\eta}_{B_\ell} \in \Xi$, $\hat{\alpha}_{jB_\ell}\left(\cdot, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right) \in \Gamma(X)$, *and* $\hat{\sigma}_j\left(\cdot, \hat{\alpha}_{jB_\ell}\left(X, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right), \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right) \in \Gamma(Z_j)$, *for all* j;

ii) *With probability approaching one, for all* k, $\left|\left|\hat{M}_{j\ell k} - M_{jk}\right|\right|_2$ *is bounded by one of the following:* $\left|\left|\hat{\theta}_{A_\ell} - \theta_0\right|\right|$, $\left|\left|\hat{\theta}_{B_\ell} - \theta_0\right|\right|$, $||\hat{\eta}_{A_\ell} - \eta_0||_\Xi$, $||\hat{\eta}_{B_\ell} - \eta_0||_\Xi$, $\left|\left|\hat{\alpha}_{jB_\ell}\left(\cdot, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right) - \alpha_{0j}(\cdot, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell})\right|\right|_2$, $\left|\left|\hat{\sigma}_{C_\ell}\left(\cdot, \hat{\alpha}_{jB_\ell}\left(X, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right), \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right) - \sigma_{0j}\left(\cdot, \hat{\alpha}_{jB_\ell}\left(X, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}\right), \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right)\right|\right|_2$. *Additionally, these are all bounded by* Cn^{-d_j}, *where* $0 < d_j < 1/2$, $j = 1, \cdots, J$, *and* C *possibly depends on* j.

Asymptotic results of OR-IVs X

- Assumption 9 implies then, for any k,

$$\left|\left|\hat{M}_{j\ell k} - M_{jk}\right|\right|_2 = \left|\left|\hat{\sigma}_{jC_\ell}\left(\cdot, \hat{\alpha}_{jB_\ell}\left(\cdot, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}\right)\right) - \sigma_0\left(\cdot, \alpha_{0j}\left(\cdot, \theta_0, \eta_0, \theta_0, \eta_0\right)\right)\right|\right|_2$$
$$= O_p\left(n^{-d_j}\right).$$

- Therefore, $\varepsilon_{jn}^M = n^{-d_j}$ in Assumption 4. By a boundness condition, we can also determine a specific rate for $\hat{F}_{j\ell}$.

Asymptotic results of OR-IVs XI

Lemma

Assume there are constants $c_1, \cdots, c_J$ such that, with probability approaching one,

$$\max_{1 \le k \le r} \left| \hat{M}_{jk}(Z_j) \right| \le c_j, \quad a.s.,$$

Additionally, let Assumption 9 hold and $r < C_1 n^{c_2}$, for some $C_1 > 0$ and $c_2 > 1$. Then,

$$\left\| \hat{F}_{j\ell} - F_j \right\|_\infty = O_p\left(n^{-d_j}\right).$$

Back

Asymptotic Results of DGMM I

- We start by imposing some regularity conditions:

Assumption

For all j and ℓ, $\mathbb{E}\left[||\psi\left(W, \theta_0, \eta_0, \boldsymbol{\kappa_0}\right)||^2 \right]$, $\left|\left|\boldsymbol{\kappa_{0j}(Z_j)}\right|\right| < \infty$ *a.s., and*

i) $\int \left| m_j\left(y, \theta_0, \hat{\eta}_\ell\right) - m_j\left(y, \theta_0, \eta_0\right)\right|^2 F_0\left(dw\right) \xrightarrow{p} 0$;

ii) $\int \left| m_j\left(y, \theta_0, \hat{\eta}_\ell\right) - m_j\left(y, \theta_0, \eta_0\right)\right|^2 \left|\left|\boldsymbol{\kappa_{0j}(z_j)}\right|\right|^2 F_0\left(dw\right) \xrightarrow{p} 0$;

iii) $\int \left| m_j\left(y, \theta_0, \eta_0\right)\right|^2 \left|\left|\boldsymbol{\hat{\kappa}_{j\ell}(z_j)} - \boldsymbol{\kappa_{0j}(z_j)}\right|\right|^2 F_0\left(dw\right) \xrightarrow{p} 0$.

Assumption

For all j and ℓ, with probability approaching one, $\hat{M}_{j\ell}\left(Z_{ji}\right)' \hat{\beta}_\ell$ *belongs to* $L^2\left(Z_j\right)$, $\left|\left|\boldsymbol{\hat{\kappa}_{j\ell}(Z_j)}\right|\right| < \infty$ *a.s., and* $\hat{\eta}_\ell \in \Xi$.

Asymptotic Results of DGMM II

Assumption

i) $||\hat{\eta}_\ell - \eta_0||_\Xi = O_p\left(n^{-d_\eta}\right)$; *ii)* $\frac{2\xi d + d_\eta(2\xi+1)}{2\xi+1} > \frac{1}{2}$, *where* $d = \min_j d_j$, $1/4 < d_\eta < 1/2$ *and* $1/4 < d < 1/2$.

Assumption

$\overline{\psi}\left(\theta_0, \eta, \boldsymbol{\kappa_0}\right)$ *is twice continuously Fréchet differentiable in a neighborhood of* η_0.

Notice:

$$\overline{\psi}\left(\theta_0, \eta, \boldsymbol{\kappa_0}\right) = \mathbb{E}\left[\psi\left(W, \theta_0, \eta, \boldsymbol{\kappa_0}\right)\right]$$

Asymptotic Results of DGMM III

- All the previous conditions yield the most important result of this section:

Lemma

Let Assumptions 1, 10-13 hold. In addition, let the assumptions in Lemma 5 hold. Then,

$$\sqrt{n}\hat{\psi}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(W_i, \theta_0, \eta_0, \boldsymbol{\kappa_0}\right) + o_p(1). \tag{16}$$

Assumption

$$\int \left| m_j\left(y, \tilde{\theta}_\ell, \hat{\eta}_\ell\right) - m_j\left(y, \theta_0, \hat{\eta}_\ell\right) \right|^2 \left|\left| \boldsymbol{\hat{\kappa}_{j\ell}(z_j)} \right|\right|^2 F_0(dw) \xrightarrow{p} 0.$$

Asymptotic Results of DGMM IV

Assumption

Υ *exists and there is a neighborhood* $\mathcal{N}$ *of* θ_0 *such that for all* ℓ

i) $||\hat{\eta}_\ell - \eta_0||_\Xi \int ||\hat{\boldsymbol{\kappa}}_{\boldsymbol{\ell}}(\boldsymbol{z}) - \boldsymbol{\kappa}_{\mathbf{0}}(\boldsymbol{z})||_\infty^2 \, dF_0(dz) \xrightarrow{p} 0;$

ii) *For* $||\eta - \eta_0||_\Xi \mathbb{E}\left[||\boldsymbol{\kappa}(\boldsymbol{Z}) - \boldsymbol{\kappa}_{\mathbf{0}}(\boldsymbol{Z})||_\infty^2\right]$ *small enough,* $\psi(W, \theta, \eta, \boldsymbol{\kappa})$ *is differentiable in* θ *on* $\mathcal{N}$ *and there are* $C_1 > 0$*,* $C_2 > 0$*, and* $d(W, \eta, \boldsymbol{\kappa})$ *such that for* $\theta \in \mathcal{N}$

$$\left|\left|\frac{\partial \psi(W, \theta, \eta, \boldsymbol{\kappa})}{\partial \theta} - \frac{\partial \psi(W, \theta_0, \eta, \boldsymbol{\kappa})}{\partial \theta}\right|\right| \leq d(W, \eta, \boldsymbol{\kappa}) ||\theta - \theta_0||^{C_1}; \quad \mathbb{E}[d(W, \eta, \boldsymbol{\kappa})] < C_2;$$

iii) *For each* s *and* k*,* $\int \left|\frac{\partial \psi_s(w, \theta_0, \hat{\eta}_\ell, \hat{\boldsymbol{\kappa}}_{\boldsymbol{\ell}})}{\partial \theta_k} - \frac{\partial \psi_s(w, \theta_0, \eta_0, \boldsymbol{\kappa}_{\mathbf{0}})}{\partial \theta_k}\right| F_0(dw) \xrightarrow{p} 0.$

Asymptotic Results of DGMM V

Theorem

Let Assumptions 1, 10-14 hold. In addition, let the assumptions in Lemma 5 hold. Also, let $\hat{\theta} \xrightarrow{p} \theta_0$, $\hat{\Lambda} \xrightarrow{p} \Lambda$, *and* $\Upsilon' \Lambda \Upsilon$ *be non-singular. Then,*

$$\sqrt{n}\left(\hat{\theta} - \theta_0\right) \xrightarrow{d} N(0, V), \quad V = \left(\Upsilon' \Lambda \Upsilon\right)^{-1} \Upsilon' \Lambda \Psi \Lambda \Upsilon \left(\Upsilon' \Lambda \Upsilon\right)^{-1}.$$

If Assumption 15 also holds, then $\hat{V} \xrightarrow{p} V$.

Asymptotic Results of DGMM VI

Theorem

If i) $\hat{\Lambda} \xrightarrow{p} \Lambda$*, where* Λ *is a positive definite matrix; ii)* $\mathbb{E}\left[\psi\left(W, \theta, \eta_0, \boldsymbol{\kappa_0}\right)\right] = 0$ *if and only if* $\theta = \theta_0$*; iii)* Θ *is compact; iv) For all* j *and* ℓ*,* $\int \left|\left|m_j\left(y, \theta, \hat{\eta}_\ell\right) \boldsymbol{\hat{\kappa}_{j\ell}(z_j)} - m_j\left(y, \theta, \eta_0\right) \boldsymbol{\kappa_{0j}(z_j)}\right|\right| F_0(dw) \xrightarrow{p} 0$ *and* $\mathbb{E}\left[\left|\left|m_j\left(Y, \theta, \eta_0\right) \boldsymbol{\kappa_{0j}(Z_j)}\right|\right|\right] < \infty$ *for all* $\theta \in \Theta$*; v) There is* $C_1 > 0$*,* $C_2 > 0$*, and* $d\left(W, \eta, \boldsymbol{\kappa}\right)$ *such that for each* $||\eta - \eta_0||_{\Xi} \mathbb{E}\left[||\boldsymbol{\kappa}(\boldsymbol{Z}) - \boldsymbol{\kappa_0}(\boldsymbol{Z})||_\infty^2\right]$ *small enough and all* $\tilde{\theta}$*,* $\theta \in \Theta$*,*

$$\left|\left|\psi\left(W, \tilde{\theta}, \eta, \boldsymbol{\kappa}\right) - \psi\left(W, \theta, \eta, \boldsymbol{\kappa}\right)\right|\right| \leq d\left(W, \eta, \boldsymbol{\kappa}\right) \left|\left|\tilde{\theta} - \theta\right|\right|^{C_1}, \quad \mathbb{E}\left[d\left(W, \eta, \boldsymbol{\kappa}\right)\right] < C_2.$$

Then, $\hat{\theta} \xrightarrow{p} \theta$*.*

Back

Additional Monte Carlo Details I

- To increase the reliability of the results, I have reduced the dimension of the problem such that I see θ_{01} and $\theta_{0\omega}$ as functions of θ_{0k}.
  - I only search over the dimension θ_{0k}.
- Notice

$$\eta_{0t}(Z_t) = \theta_{01} + \theta_{0k} K_t + \omega_t(I_t, K_t),$$

- which implies that

$$\theta_{01} + \omega_t(I_t, K_t) = \eta_{0t}(Z_t) - \theta_{0k} K_t. \tag{17}$$

- As ω_t follows an AR(1) process, I have

$$\omega_t = \theta_{0\omega}\omega_{t-1} + \epsilon_t^{\omega}, \qquad \mathbb{E}\left[\epsilon_t^{\omega} \middle| \omega_{t-1}\right] = 0. \tag{18}$$

Additional Monte Carlo Details II

- Plugging (17) into (18) and re-arranging terms yields

$$\eta_{0t}\left(Z_t\right) - \theta_{0k} K_t = \tilde{c} + \theta_{0\omega}\left(\eta_{0,t-1}\left(Z_{t-1}\right) - \theta_{0k} K_{t-1}\right) + \epsilon_t^{\omega}, \quad \tilde{c} = \theta_{01}\left(1 - \theta_{0\omega}\right).$$

- Hence, for a given value of θ_{0k}, I can identify $\theta_{0\omega}$ as the slope in a linear regression of $\eta_{0t} - \theta_{0k} K_t$ on $\eta_{0,t-1} - \theta_{0k} K_{t-1}$.
- The parameter θ_{01} can also be identified from this regression equation by using the equality $\theta_{01} = \tilde{c}/(1 - \theta_{0\omega})$, provided that $\theta_{0\omega} \neq 1$.
- As $\theta_{01} = 0$ in the Monte Carlo experiments, I directly consider $\tilde{c} = \theta_{01}$.

Additional Monte Carlo Details III

- Then, in the non-linear search, I impose these restrictions and minimize the GMM objective function based on ψ, treating it as a function of θ_{0k} only.

Back

References I

Ackerberg, Daniel, Xiaohong Chen, Jinyong Hahn, and Zhipeng Liao (2014): "Asymptotic Efficiency of Semiparametric Two-Step GMM," *Review of Economic Studies*, 81 (3), 919–943.

Ackerberg, Daniel A, Kevin Caves, and Garth Frazer (2015): "Identification Properties of Recent Production Function Estimators," *Econometrica*, 83 (6), 2411–2451.

Ai, Chunrong and Xiaohong Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71 (6), 1795–1843.

——— (2012): "The Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions," *Journal of Econometrics*, 170 (2), 442–457.

Alvarez, Roberto and Ricardo A López (2005): "Exporting and Performance: Evidence From Chilean Plants," *Canadian Journal of Economics/Revue canadienne d'économique*, 38 (4), 1384–1400.

References II

Argañaraz, Facundo and Juan Carlos Escanciano (2023): "On the Existence and Information of Orthogonal Moments For Inference," *arXiv preprint arXiv:2303.11418*.

——— (2024): "Machine Learning Debiasing with Conditional Moment Restrictions: An Application to LATE," *arXiv preprint arXiv:2410.23785*.

Bakhitov, Edvard (2022): "Automatic Debiased Machine Learning in Presence of Endogeneity," *Working Paper, https://edbakhitov.com/assets/pdf/jmp_edbakhitov.pdf*.

Brown, Bryan W and Whitney K Newey (1998): "Efficient Semiparametric Estimation of Expectations," *Econometrica*, 66 (2), 453–464.

Carrasco, Marine and Jean-Pierre Florens (2000): "Generalization of GMM to a Continuum of Moment Conditions," *Econometric Theory*, 16 (6), 797–834.

References III

Cha, Jooyoung, Harold D. Chiang, and Yuya Sasaki (2023): "Inference in High-Dimensional Regression Models without the Exact or Lp Sparsity," *The Review of Economics and Statistics*, 1–32.

Chamberlain, Gary (1992): "Efficiency Bounds for Semiparametric Regression," *Econometrica: Journal of the Econometric Society*, 567–596.

Chen, Xiaohong (2007): "Large Sample Sieve Estimation of Semi-Nonparametric Models," *Handbook of Econometrics*, 6, 5549–5632.

Chen, Xiaohong and Andres Santos (2018): "Overidentification in Regular Models," *Econometrica*, 86 (5), 1771–1817.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins (2018): "Double/debiased Machine Learning for Treatment and Structural Parameters," *The Econometrics Journal*, 21, C1–C68.

References IV

Chernozhukov, V., J.C. Escanciano, H. Ichimura, and W.K. Newey (2016): "Locally Robust Semiparametric Estimation," *arXiv preprint arXiv:1608.00033v1*.

Chernozhukov, Victor, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins (2022a): "Locally Robust Semiparametric Estimation," *Econometrica*, 90 (4), 1501–1535.

Chernozhukov, Victor, Whitney K Newey, and Rahul Singh (2022b): "Automatic Debiased Machine Learning of Causal and Structural Effects," *Econometrica*, 90 (3), 967–1027.

Friedman, Jerome, Trevor Hastie, Holger Höfling, and Robert Tibshirani (2007): "Pathwise Coordinate Optimization," *The Annals of Applied Statistics*, 1 (2), 302 – 332.

References V

Friedman, Jerome, Trevor Hastie, and Rob Tibshirani (2010): "Regularization Paths for Generalized Linear Models Via Coordinate Descent," *Journal of Statistical Software*, 33 (1), 1.

Fu, Wenjiang J (1998): "Penalized Regressions: the Bridge Versus the Lasso," *Journal of Computational and Graphical Statistics*, 7 (3), 397–416.

Gandhi, Amit, Salvador Navarro, and David A Rivers (2020): "On the Identification of Gross Output Production Functions," *Journal of Political Economy*, 128 (8), 2973–3016.

Levinsohn, James and Amil Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables," *The Review of Economic Studies*, 70 (2), 317–341.

Luenberger, David G (1997): *Optimization By Vector Space Methods*, John Wiley & Sons.

References VI

NEYMAN, J. (1959): "Optimal Asymptotic Tests of Composite Statistical Hypothesis," *Probability and Statistics: The Harald Cramer Volume*, 213–234.

OLLEY, G. STEVEN AND ARIEL PAKES (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64 (6), 1263–1297.

PAVCNIK, NINA (2002): "Trade Liberalization, Exit, and Productivity Improvements: Evidence From Chilean Plants," *The Review of Economic Studies*, 69 (1), 245–276.

WOOLDRIDGE, JEFFREY M (2009): "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics letters*, 104 (3), 112–114.