Debiasing Structural Parameters with General Conditional Moments and High-Dimensional First Stages

Facundo Argañaraz

Universidad Carlos III de Madrid

June 3, 2024

ENTER Jamboree 2024

This paper is about I

- A method to conduct (GMM) inference on a finite-dimensional parameter.
    - Models defined by a finite number of conditional moment restrictions (CMRs).
    - Possibly different conditioning variables.
    - Endogenous regressors.
- Examples:
    - Regression, quantile, missing data, dynamic discrete choice, non-linear simultaneous equations, production functions, and many other models (see Chen and Qiu, 2016)

This paper is about II

- CMRs are allowed to depend on non-parametric components.
  - Machine Learning tools, e.g., Lasso, Boosting, Random Forest, Neural Networks,...
  - First stage bias.
  - Bias decays at a rate slower than $\sqrt{n}$.
  - Plugging-in is not a good idea.
- Inference is based on Locally Robust (LR)/Orthogonal/Debiased moments, extended to the case with CMRs.
  - Less affected by first-stage bias than non-orthogonal moments (when plugging in).
  - Standard inference is typically valid.
- A general procedure to construct those.
  - Data-driven (or automatic).

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 output of firm i at time t, and X_{it} be a vector 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}.
  - 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).
- We assume that there exists some firm's choice I_{it}, e.g., investment, at t that is linked to ω_{it}:

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

- No parametric assumptions are imposed on I_t, except for a strict monotonicity condition (in ω_t).
- We shall write

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

Example: Production Functions III

- Equation (1) becomes

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

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

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

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

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

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

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

Production Functions IV

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

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

- Our goal is to learn $\theta_0 = \left(\theta_{0p}', \theta_{0\omega}\right)'$, in the presence of an unknown η_0.

Production Functions V

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

$$\mathbb{E}\left[Y_1 - \eta_{01}(I_1, X_1) \mid I_1, X_1 \right] = 0, \tag{2}$$
$$\mathbb{E}\left[Y_2 - F(X_2, \theta_{0p}) - \theta_{0\omega}\left(\eta_{01}(I_1, X_1) - F(X_1, \theta_{0p})\right) \mid \Omega_1 \right] = 0, \tag{3}$$
$$\mathbb{E}\left[Y_2 - \eta_{02}(I_2, X_2) \mid I_2, X_2 \right] = 0, \tag{4}$$
$$\mathbb{E}\left[Y_3 - F(X_3, \theta_{0p}) - \theta_{0\omega}\left(\eta_{02}(I_2, X_2) - F(X_2, \theta_{0p})\right) \mid \Omega_2 \right] = 0. \tag{5}$$

- Estimation based on non-orthogonal moments using a plug-in procedure:
  1. Step 1: Employ, e.g., Random Forest and estimate $\eta_0 = (\eta_{01}, \eta_{02})$, using (2) and (4).
  2. Step 2: Select IVs based on Ω_t, e.g., $r(\Omega_t) = (I_t, X_t, I_{t-1}, X_{t-1})'$ and use GMM based on (3) and (5):

$$\mathbb{E}\left[\left(Y_2 - F(X_2, \theta_{0p}) - \theta_{0\omega}\left(\eta_{01}(I_1, X_1) - F(X_1, \theta_{0p})\right)\right) \otimes r(\Omega_1)\right] = 0$$
$$\mathbb{E}\left[\left(Y_3 - F(X_3, \theta_{0p}) - \theta_{0\omega}\left(\eta_{02}(I_2, X_2) - F(X_2, \theta_{0p})\right)\right) \otimes r(\Omega_2)\right] = 0.$$

- What is the distribution of $\sqrt{n}\left(\hat{\theta} - \theta_0\right)$?

Figure: Comparison of Non-Orthogonal and Orthogonal Estimators

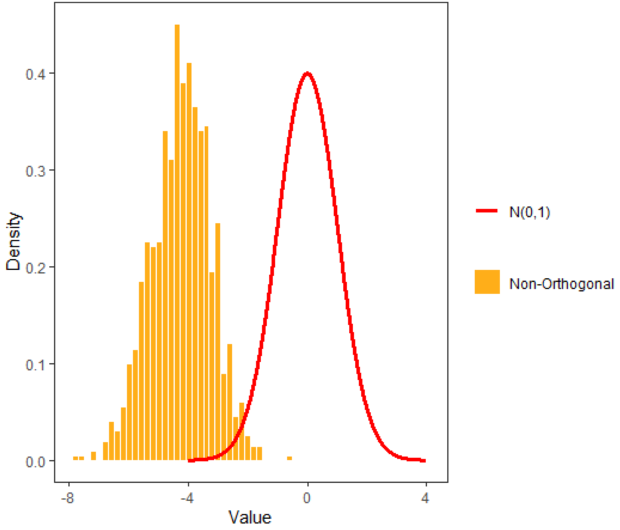
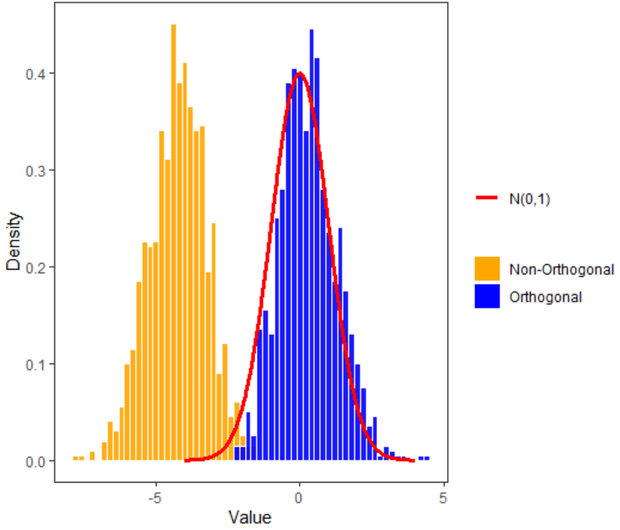


Figure: Comparison of Non-Orthogonal and Orthogonal Estimators



Debiased Moments?

Debiased Moments

- A debiased moment is a moment locally "insensitive" to η_0.

Debiased Moments

- A debiased moment is a moment locally “insensitive” to η_0.
- To achive that insensitivity we need to introduce an extra nuisance κ_0 (Chernozhukov et al., 2022a).

Debiased Moments

- A debiased moment is a moment locally "insensitive" to η_0.
- To achive that insensitivity we need to introduce an extra nuisance κ_0 (Chernozhukov et al., 2022a).
- A debiased moment in our setting is a moment based on a function $\psi : \mathcal{W} \times \Theta \times \boldsymbol{B} \times L^2(Z) \mapsto \mathbb{R}$ satisfying the following two restrictions:

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

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

Debiased Moments

- A debiased moment is a moment locally "insensitive" to η_0.
- To achive that insensitivity we need to introduce an extra nuisance κ_0 (Chernozhukov et al., 2022a).
- A debiased moment in our setting is a moment based on a function $\psi : \mathcal{W} \times \Theta \times \boldsymbol{B} \times L^2(Z) \mapsto \mathbb{R}$ satisfying the following two restrictions:

$$\begin{aligned} \frac{d}{d\tau}\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0 + \tau b, \kappa_0\right)\right] &= 0, \quad \text{for all } b \in \boldsymbol{B}, \\ \mathbb{E}\left[\psi\left(W, \theta_0, \eta_0, \kappa\right)\right] &= 0, \quad \text{for all } \kappa \in L^2(Z). \end{aligned}$$

- How can we construct ψ in our example?

Debiased Moments

- A debiased moment is a moment locally "insensitive" to η_0.
- To achive that insensitivity we need to introduce an extra nuisance κ_0 (Chernozhukov et al., 2022a).
- A debiased moment in our setting is a moment based on a function $\psi : \mathcal{W} \times \Theta \times \boldsymbol{B} \times L^2(Z) \mapsto \mathbb{R}$ satisfying the following two restrictions:

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

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

- How can we construct ψ in our example?
  - Simply combine the initial residuals functions (Argañaraz and Escanciano, 2023).

Example (continued)

- We can obtain a debiased moment by means of

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

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

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

$$\begin{aligned}&\frac{d}{d\tau}\mathbb{E}\left[\psi\left(W,\theta_0,\eta_0+\tau b,\kappa_0\right)\right] \\ &=\mathbb{E}\left[b_1\left(Z_1\right)\left(-\kappa_{01}\left(Z_1\right)-\theta_{0\omega}\kappa_{02}\left(Z_1\right)\right)+b_2\left(Z_2\right)\left(-\kappa_{02}\left(Z_2\right)-\theta_{0\omega}\kappa_{02}\left(Z_2\right)\right)\right] \\ &=0.\end{aligned}$$

How can we get κ_0?

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

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

- Technically, κ_0 is orthogonal to $\overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}$.

Estimation of OR-IVs (or κ_0's)

Estimation of OR-IVs I

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

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

- $\Pi_{\overline{\mathcal{R}(S_{\theta_0,\eta_0})}}$ denotes the orthogonal projection operator onto $\overline{\mathcal{R}(S_{\theta_0,\eta_0})}$ (or "fitted values").
- Approximate $\Pi_{\overline{\mathcal{R}(S_{\theta_0,\eta_0})}} f = f^*$.
  - A minimization problem.
  - Use the operators $S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0} g$.
  - Need to find the $g*$ such that $S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0} g*$ is close to f^* (or f).
  - Look for a solution $g \in \mathcal{G}$.
  - $S^{(j)}_{\theta_0,\eta_0} S^*_{\theta_0,\eta_0}$ is unknown $\rightarrow$ Estimate it.
  - Potentially, more than one solution exists $\rightarrow$ Focus on the minimum norm solution.

Estimation of OR-IVs II

- We propose to estimate g_0 by means of

$$\hat{g}_n = \arg\min_{g \in \mathcal{G}_n} \sum_{j=1}^{J} \mathbb{E}\left[\left(f_j(Z_j) - \hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}} g\right)^2\right] + 2\lambda_n \, ||g||^2_{\mathcal{G}} \, ,$$

- To compute $\hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}}$ use **cross-fitting**.
  - Randomly partition the sample into L subgroups, $I_1, \cdots, I_L$, of the same size.
  - Let I^c_ℓ be the complement of I_ℓ.
  - Estimate $\hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S^*}_{\hat{\theta},\hat{\eta}}$ using I^c_ℓ.
- Focus on a particular $\mathcal{G}_n$.

Estimation of OR-IVs III

- In this paper, $\mathcal{G}_n$ is the **space of sparse functions**:

$$\mathcal{G}_n = \left\{ g : g_j(Z_j) = \gamma_j(Z_j)' \beta_j, \;\; ||\beta||_0 = s, \;\; ||\beta||_\infty < c \right\}.$$

  where $\gamma(Z) = \left(\gamma_1(Z_1)', \cdots, \gamma_J(Z_J)'\right)'$ is a vector of basis functions.

- Then, we only need to focus on obtaining an optimal $\hat{\beta}$:

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

  where $\boldsymbol{\hat{M}_{j\ell}}$'s are estimated regressors.

- A **Lasso**-type program with estimated regressors.

Estimation of OR-IVs - Recap

- Let $\boldsymbol{f_{j\ell}}$ be a $n_\ell-$dimensional vector containing each $f_j(Z_{ji})$, $i \notin I_\ell$.
  - Racall: you provide me with an $f(Z)$, e.g., $f(Z) = Z$.
- Let $\boldsymbol{\hat{M}_{j\ell}}$ be a suitable $n_\ell \times r$ design matrix associated with $\hat{S}^{(j)}_{\hat{\theta},\hat{\eta}} \hat{S}^*{}_{\hat{\theta},\hat{\eta}}$.
- The estimator $\hat{\beta}_\ell$ can be written as follows ▸ More details

▸ Coordinate Descent Approach

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

- $\hat{\kappa}_\ell$ is the “residual” of the previous program.

More in the paper

More in the paper

1. A **general** setting (more details):

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

2. Some regularity conditions are sufficient to show

$$\boxed{||\hat{\kappa}(Z)-\kappa_0(Z)||_{L^2(Z)}=O_p\left(\mu_n^{\kappa}\right), \quad \mu_n^{\kappa}=\sqrt{s}\lambda_n.}$$

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

3. Introduce a GMM estimator $\hat{\theta}$ for θ_0 in a Two-Step setting.
   - More details.
4. Some regularity conditions are sufficient to show

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

5. $\hat{V}\xrightarrow{p} V$.

Monte Carlo

Monte Carlo I

More details

- Example.
- Firms are followed during three periods, i.e., $T = 3$.
- Cobb-Douglass 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

- This decision is assumed to follow

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

- where $\gamma_0 = 0$, $\gamma_1 = -0.7$, and $\gamma_2 = 5$.
- Productivity is assumed to follow a normal AR(1) process with $\theta_{0\omega} = 0.7$.

Monte Carlo III

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

$$
\begin{aligned}
f_1(Z) &= (K_{i1}, K_{i1}, K_{i2}, K_{i2})', \\
f_2(Z) &= (I_{i1}, I_{i1}, I_{i2}, I_{i2})', \\
f_3(Z) &= (K_{i1}, K_{i1}, I_{i2}, I_{i2})', \\
f_4(Z) &= (K_{i1}, I_{i1}, I_{i2}, I_{i2})'.
\end{aligned}
$$

- These are choices that people use in applied work to estimate θ_0 by GMM, but they lead to non-orthogonal moments.

Monte Carlo IV

- In all situations, the bases coincide, i.e., $\gamma_j = \tilde{\gamma}$, and β_j's are assumed to be constant across j, for simplicity.
- η_0 is estimated with Boosting.
- $L = 4$.
- γ's are exponential bases.
- $r = 9$ (recall $\beta \in \mathbb{R}^r$).
- $\lambda_n = \frac{1.1}{\sqrt{n-n_\ell}}\Phi^{-1}\left(1 - \frac{c_2}{2r}\right)$, with $c_2 = 0.1/\log((n - n_\ell) \vee r)$ (Belloni et al., 2012, BCCH).

Figure: Monte Carlo Results - Bias and 95% Coverage

			$n = 250$				
Est.	Smaller λ_n	Larger λ_n	λ_n (BCCH)	Larger L	Larger r	Fourier Basis	Random Forest
Bias ($\hat{\theta}_1$)	0.095	0.097	0.100	0.105	0.095	0.105	0.100
Cov95%	0.935	0.934	0.936	0.912	0.937	0.948	0.914
Bias ($\hat{\theta}_k$)	-0.031	-0.039	-0.041	-0.044	-0.036	-0.046	-0.042
Cov95%	0.912	0.913	0.906	0.894	0.910	0.925	0.918
Bias ($\hat{\theta}_\omega$)	-0.160	-0.162	-0.163	-0.165	-0.160	-0.166	-0.253
Cov95%	0.738	0.742	0.739	0.651	0.745	0.733	0.777

Figure: Monte Carlo Results - Bias and 95% Coverage (continued)

	$n = 500$						
Est.	Smaller λ_n	Larger λ_n	λ_n (BCCH)	Larger L	Larger r	Fourier Basis	Random Forest
Bias ($\hat{\theta}_1$)	0.048	0.061	0.059	0.060	0.059	0.071	0.035
Cov95%	0.943	0.939	0.947	0.927	0.941	0.959	0.963
Bias ($\hat{\theta}_k$)	-0.013	-0.029	-0.027	-0.027	-0.027	-0.040	-0.021
Cov95%	0.903	0.935	0.927	0.894	0.935	0.935	0.949
Bias ($\hat{\theta}_\omega$)	-0.081	-0.088	-0.087	-0.074	-0.087	-0.095	-0.103
Cov95%	0.926	0.922	0.922	0.886	0.922	0.919	0.970

Figure: Monte Carlo Results - Bias and 95% Coverage (continued)

	$n = 750$						
Est.	Smaller λ_n	Larger λ_n	λ_n (BCCH)	Larger L	Larger r	Fourier Basis	Random Forest
Bias ($\hat{\theta}_1$)	0.028	0.039	0.037	0.038	0.039	0.053	0.022
Cov95%	0.944	0.946	0.949	0.955	0.958	0.965	0.980
Bias ($\hat{\theta}_k$)	-0.002	-0.020	-0.017	-0.017	-0.020	-0.037	-0.018
Cov95%	0.880	0.929	0.925	0.924	0.930	0.944	0.945
Bias ($\hat{\theta}_\omega$)	-0.018	-0.025	-0.023	-0.012	-0.025	-0.033	-0.041
Cov95%	0.952	0.951	0.954	0.952	0.951	0.950	0.990

Final Remarks

- Our approach will hopefully pave the way for the employment of machine learning techniques in context where the construction of LR has remained unexplored.
- In future versions, we plan to use data from a panel of Chilean firms.
  - This data has been extensively studied by the production function literature; see, e.g., Levinsohn and Petrin (2003), Ackerberg et al. (2015), and Gandhi et al. (2020).
  - Can our strategy uncover larger heterogeneity patterns among production functions than previously recognized?
- In subsequent works...
  - Identification and efficiency (or other notions of optimality (?)).
  - A general framework for different $\mathcal{G}_n$'s.
  - More general parameters.

APPENDIX

Algorithm to estimate OR-IVs I

- **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 $\gamma(Z)$.[1]
- **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 $\hat{\mathbb{E}}_{C_\ell}[\cdot | Z_j]$. These conditional expectations depend on known $\tilde{\nu}_j$, and thus can be evaluated.
- **Step 2:** Compute design matrix $\hat{\boldsymbol{M}}_{\boldsymbol{j\ell}}$ such that its $(i,l)-$entry is

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

Algorithm to estimate OR-IVs II

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

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

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

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

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

$$\hat{\epsilon}_{ji\ell} = f_j\left(Z_{ji}\right) - \sum_{j'=1}^J\sum_{k=1}^{r_{j'}}\hat{\beta}_{j'k\ell}\hat{\mathbb{E}}_{C_\ell}\left[\left(\hat{\mathbb{E}}_{B_\ell}\left[\tilde{\nu}_{j'}\left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{jA_\ell}\right)\gamma_{j'k}\left(Z_{j'i}\right)\middle| X\right]\right)' \tilde{\nu}_j\left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{jB_\ell}\right)\middle| Z_{ji}\right].$$

Algorithm to estimate OR-IVs III

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

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

and

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

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

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

$$\begin{aligned} \hat{\kappa}_{j\ell}(Z_{ji}) &= f_j(Z_j) - \hat{f_j^*}(Z_j) \\ &= f_j(Z_{ji}) - \sum_{j'=1}^{J} \sum_{k=1}^{r_{j'}} \hat{\beta}_{j'k\ell} \hat{\mathbb{E}}_{C_\ell} \left[\left(\hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'} \left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell} \right) \gamma_{j'k}(Z_{ji}) \middle| X \right] \right)' \tilde{\nu}_j \left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell} \right) \middle| Z_{ji} \right]. \end{aligned} \tag{6}$$

▸ Back

[1] E.g., take the first $\tilde{r}_j$ components of each γ_j.

Coordinate Descent Approach I

- Step 4 of the iterative algorithm above requires to solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^{J} \frac{1}{n - n_\ell} \left(\boldsymbol{f_{j\ell}} - \hat{\boldsymbol{M}}_{j\ell}\beta \right)' \left(\boldsymbol{f_{j\ell}} - \hat{\boldsymbol{M}}_{j\ell}\beta \right) + 2\lambda_n \left\| \hat{D}_\ell \beta \right\|_1, \quad (7)$$

- 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 (7), we 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, we 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):

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

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

Coordinate Descent Approach III

2 **Step 2:** Update coordinate β_l:

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

Back

General Setting I

- The data $W_i = (Y_i, X_i, Z_i)$, $i = 1, \cdots, n$, is iid.
- Let $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ denote a finite-dimensional parameter vector.
- Let $\eta \in \boldsymbol{B}$ be a vector of real-valued measurable functions of X.
- To be specific, $\eta = \left(\eta_1, \cdots, \eta_{d_\eta}\right)$ with $\eta_s \equiv \eta_s(X)$.
- There is a vector of residual functions $m_j : \mathcal{Y} \times \Theta \times \boldsymbol{B} \mapsto \mathbb{R}$ such that:

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

- m_j might depend on θ_0 arbitrarily.
- There exists a unique $(\theta_0, \eta_0) \in \Theta \times \boldsymbol{B}$ such that (40) holds.
- Let $\kappa = (\kappa_1, \cdots, \kappa_J)$, where $\kappa_j \equiv \kappa_j(Z_j)$, and $\kappa_j \in L^2(Z_j)$.

General Setting II

- Let $\boldsymbol{B} \subseteq \bigotimes^{d_\eta} L^2(X)$ be a Hilbert space and define

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

Assumption

Given some $||\cdot||$, $h_j(Z_j, \theta_0, \cdot) : \boldsymbol{B} \mapsto L^2(Z_j)$ is Fréchet differentiable in a neighborhood of η_0, where the derivative is given by

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

for some $b \in \boldsymbol{B}$.

General Setting III

- Remark that (1) defines a linear operator $S^{(j)}_{\theta_0,\eta_0} : \boldsymbol{B} \mapsto L^2(Z_j)$. In addition, let us define

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

- $S_{\theta_0,\eta_0} : \boldsymbol{B} \mapsto L^2(Z)$ is also a linear operator.
- S_{θ_0,η_0} simply "collects" all the possible derivatives of the CMRs with respect to η_0.
- It is sufficient to find κ_0 orthogonal to such a collection.
- In formal terms, κ_0 needs to be orthogonal to the range of S_{θ_0,η_0}.

General Setting IV

- The range of S_{θ_0,η_0} is given by

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

- A key object:

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

- Let $\kappa_0 \in \overline{\mathcal{R}\left(S_{\theta_0,\eta_0}\right)}^{\perp}$.

- Then, it can be easily verified that a debiased moment can be constructed as follows:

$$\psi\left(W, \theta_0, \eta_0\right) = \sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0\right) \kappa_{0j}\left(Z_j\right).$$

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 $M_j(Z_{ji})$.
- We define

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

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

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

$$\hat{\beta}_\ell = \underset{\beta\in\mathbb{R}^r}{\arg\min}\ \sum_{j=1}^{J}\left(-2\hat{F}_{j\ell}'\beta - \beta'\,\hat{G}_{j\ell}\beta\right) + 2\lambda_n\,||\beta||_1\,. \tag{8}$$

Asymptotic results of OR-IVs II

Assumption

There are constants $c_1, \cdots, c_J$ *such that with probability approaching one*

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

Assumption

$$r^2 \int \left\| \hat{M}_{j\ell}(z_{ji}) \hat{M}_{j\ell}(z_{ji})' - M_{j\ell}(z_{ji}) M_{j\ell}(z_{ji})' \right\|_\infty F_0(dw) = o_p\left(\varepsilon_n^2\right),$$

where $\varepsilon_n = \sqrt{\frac{\log(r)}{n}}$.

Asymptotic results of OR-IVs III

Assumption

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

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

Assumption

The largest eigenvalue of $\sum_{j=1}^{J} G_j$ is uniformly bounded in n and there is a $c > 0$ such that with probability approaching one

$$\phi^2(s) = \inf\left\{\frac{\delta' \sum_j^J \hat{G}_j \delta}{\left|\left|\delta_{S_\beta}\right|\right|_2^2}, \quad \delta \in \mathbb{R}^r \setminus \{0\}, \left|\left|\delta_{S_\beta^c}\right|\right|_1 \leq 3\left|\left|\delta_{S_\beta}\right|\right|_1, \quad |S_\beta| \leq s\right\}$$
$$> c.$$

Asymptotic results of OR-IVs IV

Assumption

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

Assumption

Let

$$B = \sum_{j=1}^{J} \int \left(M_j\left(z_j\right) - \hat{M}_j\left(z_j\right)\right)\left(M_j\left(z_j\right) - \hat{M}_j\left(z_j\right)\right)' F_0\left(dw\right).$$

Then, the maximum eigenvalue of B is $O_p\left(\varepsilon_n^2\right)$.

Asymptotic results of OR-IVs V

Theorem

Let the previous assumptions hold. In addition, suppose that $\varepsilon_n = o(\lambda_n)$*. Then,*

$$||\hat{\kappa}(Z) - \kappa_0(Z)||_{L^2(Z)} = O_p(\mu_n^\kappa), \quad \mu_n^\kappa = \sqrt{s}\lambda_n.$$

Back

Estimation of the Parameter of Interest I

- Simplify some aspects of our general model.
- Two-step setting.
  - There are functions m_j's that depend on η_0 only.
  - Many relevant scenarios in applied work present this feature (see, e.g., Chen and Qiu, 2016, Section 5 and references therein).
- Focus on the case where m_j depends on η_j only and η_{0j} is a conditional expectation.
- Notice that for different choices of instruments, say q of them, we can construct J vectors $\boldsymbol{\kappa_{0j}}(\boldsymbol{Z_j})$, of dimension q.

Estimation of the Parameter of Interest II

- Let
$$\psi\left(W,\theta,\eta,\boldsymbol{\kappa}\right)=\sum_{j=1}^{J}m_j\left(Y_i,\theta,\eta_j\right)\boldsymbol{\kappa_j(Z_j)},$$
- Let $\hat{\eta}_\ell$ be an estimator of η_0, using observations in I_ℓ^c.
- Let
$$\hat{\psi}\left(\theta\right)=\frac{1}{n}\sum_{\ell=1}^{L}\sum_{i\in I_\ell}\psi\left(W_i,\theta,\hat{\eta}_\ell,\hat{\boldsymbol{\kappa}}_\ell\right).$$
- Our proposed estimator $\hat{\theta}$ is defined as the solution to the GMM program
$$\hat{\theta}=\underset{\theta\in\Theta}{\arg\min}\ \hat{\psi}\left(\theta\right)'\hat{\Lambda}\hat{\psi}\left(\theta\right),\tag{9}$$

Estimation of the Parameter of Interest III

- A choice that asymptotically minimizes the asymptotic variance is $\hat{\Lambda} = \hat{\Psi}^{-1}$, where

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

- The estimator of the asymptotic variance, which accounts for the estimation of η_0 and κ_0, takes the "sandwich" form

$$\hat{V} = \left(\hat{\Upsilon}' \hat{\Lambda} \hat{\Upsilon} \right)^{-1} \hat{\Upsilon}' \hat{\Lambda} \hat{\Psi} \hat{\Lambda} \hat{\Upsilon} \left(\hat{\Upsilon}' \hat{\Lambda} \hat{\Upsilon} \right)^{-1}, \quad \hat{\Upsilon} = \frac{\partial}{\partial \theta} \hat{\psi}(\hat{\theta}). \tag{10}$$

Estimation of η_0

- We allow for a η_0 that depends on variables different from Z.
  - An ill-posed problem (Newey and Powell, 2003).
  - Let $T_j : L^2(X) \mapsto L^2(Z_j)$ denote the conditional expectation operator given by

$$T_j \eta_j = \mathbb{E}\left[\eta_j(X) \mid Z_j\right].$$

  - Consider the projected mean square norm:

$$||T_j(\eta_j - \eta_{0j})||_2 = \sqrt{\mathbb{E}\left[\mathbb{E}\left[\eta_j(X) - \eta_{0j}(X) \mid Z_j\right]^2\right]},$$

$$||T(\eta - \eta_0)||_{L^2(Z)} \equiv \sqrt{\sum_{j=1}^{J} ||T_j(\eta_j - \eta_{0j})||_2^2}.$$

Back

Asymptotic Results of D-CMRs I

Assumption

$$\mathbb{E}\left[||\psi\left(W,\theta_0,\eta_0,\boldsymbol{\kappa_0}\right)||^2\right] < \infty, \textit{ and}$$

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

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

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

- Let us define

$$\hat{\Delta}_\ell(w) = \sum_{j=1}^{J}\left(m_j\left(y,\theta_0,\hat{\eta}_{j\ell}\right) - m_j\left(y,\theta_0,\eta_{0j}\right)\right)\left(\boldsymbol{\hat{\kappa}_{j\ell}(Z_j)} - \boldsymbol{\kappa_{0j}(Z_j)}\right).$$

Asymptotic Results of D-CMRs II

Assumption

There are constants $c_1, \cdots, c_j$ such that with probability approaching one

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

Assumption

i) $||T(\hat{\eta}_\ell - \eta_0)||_{L^2(Z)} = O_p(\mu_n^\eta), \quad \mu_n^\eta = o\left(n^{-1/4}\right)$; *ii)* $\sqrt{n}\mu_n^\eta \mu_n^\kappa \to 0.$

Asymptotic Results of D-CMRs III

Assumption

For $||T(\hat{\eta}_\ell - \eta_0)||^2_{L^2(Z)}$ *small enough,*

$$\sum_{j=1}^{J} ||T_j(m_j(y,\theta_0,\eta_j) - m_j(y,\theta_0,\eta_{0j}))||_2^2 \leq C\,||T(\hat{\eta}_\ell - \eta_0)||^2_{L^2(Z)}.$$

- The previous assumptions and $\varepsilon_n = o(\lambda_n)$ imply

$$i)\ \int \left|\left|\hat{\Delta}_\ell(w)\right|\right|^2 F_0(dw) \xrightarrow{p} 0, \quad and \quad ii)\ \sqrt{n}\int \hat{\Delta}_\ell(w) F_0(dw) \xrightarrow{p} 0. \tag{11}$$

Asymptotic Results of D-CMRs IV

- Let

$$\overline{\psi}(\theta, \eta, \boldsymbol{\kappa}) = \mathbb{E}[\psi(W, \theta, \eta, \boldsymbol{\kappa})].$$

Assumption

$\overline{\psi}(\theta, \eta, \boldsymbol{\kappa})$ is twice continuously Fréchet differentiable in a neighborhood of η_0.

- Then it can be shown that since ψ leads to a debiased moment, there exists a $C > 0$ such that

$$\left|\left|\overline{\psi}(\theta_0, \eta, \boldsymbol{\kappa_0})\right|\right| \leq C \left|\left|T(\hat{\eta}_\ell - \eta_0)\right|\right|^2_{L^2(Z)}.$$

Asymptotic Results of D-CMRs V

- All the previous conditions are sufficient to show

$$\sqrt{n}\hat{\psi}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi\left(W_i, \theta_0, \eta_0, \boldsymbol{\kappa_0}\right) + o_p(1). \tag{12}$$

- The result in (12) is essential for obtaining asymptotic normality of $\hat{\theta}$.
- Interestingly, cross-fitting enables to show (12) in a simple manner, without the need to impose the so-called Donsker conditions for η_0, as discussed in Chernozhukov et al. (2018) and Chernozhukov et al. (2022a).

Assumption

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

Asymptotic Results of D-CMRs VI

- We need conditions for convergence of the Jacobian: $\frac{\partial}{\partial\theta}\hat{\psi}(\bar{\theta}) \xrightarrow{p} \Upsilon = \mathbb{E}\left[\frac{\partial}{\partial\theta}\psi\left(W, \theta_0, \eta_0, \boldsymbol{\kappa_0}\right)\right]$ for any $\bar{\theta} \xrightarrow{p} \theta_0$. To that end, we impose the following:

Asymptotic Results of D-CMRs VII

Assumption

Υ *exists and there is a neighborhood* $\mathcal{N}$ *of* θ_0 *and* $||\cdot||$ *such that*

i) $||T(\hat{\eta}_\ell - \eta_0)||_{L^2(Z)} \, ||\hat{\kappa}_\ell - \kappa_0||_{L^2(Z)} \xrightarrow{p} 0$;

ii) *For all* $||T(\eta - \eta_0)||_{L^2(Z)} \, ||\kappa - \kappa_0||_{L^2(Z)}$ *(where we are considering each element of* $\boldsymbol{\kappa}_j$*) small enough,* $\psi(W, \theta, \eta, \boldsymbol{\kappa})$ *is differentiable in* θ *on* $\mathcal{N}$ *with probability approaching one and there is a* C *and* $d(W, \eta, \boldsymbol{\kappa})$ *such that for* $\theta \in \mathcal{N}$ *and for each* $||T(\eta - \eta_0)||_{L^2(Z)} \, ||\kappa - \kappa_0||_{L^2(Z)}$ *small enough*

$$\left|\left|\frac{\partial \psi(W, \theta, \eta, \boldsymbol{\kappa})}{\partial \theta} - \frac{\partial \psi(W, \theta_0, \eta, \boldsymbol{\kappa})}{\partial \theta}\right|\right| \leq d(W, \eta, \boldsymbol{\kappa}) \, ||\theta - \theta_0||^{1/C}; \quad \mathbb{E}[d(W, \eta, \kappa)] < C;$$

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

Asymptotic Results of D-CMRs VIII

Theorem

Let the previous assumptions hold. In addition, 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 14 also holds, then $\hat{V} \xrightarrow{p} V$.

- Note that Theorem 2 relies on the consistency of $\hat{\theta}$.

Asymptotic Results of D-CMRs IX

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)* $\int \left|\left| m_j\left(y, \theta, \hat{\eta}_{j\ell}\right) \hat{\boldsymbol{\kappa}}_{\boldsymbol{j\ell}}(\boldsymbol{z_j}) - m_j\left(y, \theta, \eta_{0j}\right) \boldsymbol{\kappa}_{\boldsymbol{0j}}(\boldsymbol{z_j})\right|\right| F_0(dw) \xrightarrow{p} 0$ *and* $\mathbb{E}\left[\left|\left| m_j\left(Y, \theta, \eta_0\right) \boldsymbol{\kappa}_{\boldsymbol{0j}}(\boldsymbol{Z_j})\right|\right|\right] < \infty$ *for all* $\theta \in \Theta$*; v) There is a* $C > 0$ *and* $d\left(W, \eta, \boldsymbol{\kappa}\right)$ *such that for each* $\left|\left| T\left(\eta - \eta_0\right)\right|\right|_{L^2(Z)} \left|\left|\kappa - \kappa_0\right|\right|_{L^2(Z)}$ *small enough and all* $\tilde{\theta}$*,* $\theta \in \Theta$*,*

$$\left|\left|\psi\left(W, \tilde{\theta}, \eta, \boldsymbol{\kappa}\right) - \psi\left(W, \theta, \eta, \boldsymbol{\kappa}\right)\right|\right| \leq d\left(W, \eta, \boldsymbol{\kappa}\right) \left|\left|\tilde{\theta} - \theta\right|\right|^{1/C}, \quad \mathbb{E}\left[d\left(W, \eta, \boldsymbol{\kappa}\right)\right] < C.$$

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

▸ Back

Additional Monte Carlo Details I

- In our Monte Carlo experiments, we have considered different other choices:
  1. The smaller λ_n is such that $\lambda_n = \frac{1.01}{\sqrt{n-n_\ell}}\Phi^{-1}\left(1-\frac{c_2}{2r}\right)$, with $c_2 = 2/\log(\log(\log((n-n_\ell)\vee r)))$.
  2. The case with larger λ_n has $\lambda_n = \frac{1.3}{\sqrt{n-n_\ell}}\Phi^{-1}\left(1-\frac{c_2}{2r}\right)$, with $c_2 = 0.1/\log((n-n_\ell)\vee r)$.
  3. We also consider a scenario where $L = 6$.
  4. In a different experiment, we specify a larger number of coefficients such that $r = 25$.
  5. Additionally, we model γ's through Fourier basis.
  6. Finally, in another situation, η_0 is estimated with Random Forest.

Additional Monte Carlo Details II

- To obtain our estimator $\hat{\theta} = \left(\hat{\theta}_1, \hat{\theta}_k, \hat{\theta}_\omega\right)'$, we use GMM based on four debiased moments.
- These can be written as

$$\psi\left(W, \theta_0, \eta_0\right) = \left(Y_1 - \eta_{01}\left(I_1, K_1\right)\right) \boldsymbol{\kappa_{01}}\left(\boldsymbol{Z_1}\right) + \left(Y_2 - \theta_{01} - \theta_{0k} K_2 - \theta_{0\omega}\left(\eta_{01}\left(Z_1\right) - \theta_{01} - \theta_{0k} K_1\right)\right) \boldsymbol{\kappa_{02}}\left(\boldsymbol{Z_1}\right)$$
$$+ \left(Y_2 - \eta_{02}\left(I_2, K_2\right)\right) \boldsymbol{\kappa_{03}}\left(\boldsymbol{Z_2}\right) + \left(Y_3 - \theta_{01} - \theta_{0k} K_3 - \theta_{0\omega}\left(\eta_{02}\left(Z_2\right) - \theta_{01} - \theta_{0k} K_2\right)\right) \boldsymbol{\kappa_{04}}\left(\boldsymbol{Z_2}\right).$$

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

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

Additional Monte Carlo Details III

- which implies that

$$\theta_{01} + \omega_t \left(I_t, K_t \right) = \eta_{0t} \left(Z_t \right) - \theta_{0k} K_t. \tag{13}$$

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

$$\omega_t = \theta_{0\omega} \omega_{t-1} + \epsilon_t^{\omega}, \qquad \mathbb{E} \left[\epsilon_t^{\omega} \middle| \omega_{t-1} \right] = 0. \tag{14}$$

- Plugging (13) into (14) 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}, we can identify $\theta_{0\omega}$ as the slope in a linear regression of $\eta_{0t} - \theta_{0k} K_t$ on $\eta_{0,t-1} - \theta_{0k} K_{t-1}$.

Additional Monte Carlo Details IV

- 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 our Monte Carlo experiments, we directly consider $\tilde{c} = \theta_{01}$.
- Then, in our non-linear search, we 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, D., Chen, X., Hahn, J., and Liao, Z. (2014). Asymptotic Efficiency of Semiparametric Two-Step GMM. *Review of Economic Studies*, 81(3):919–943.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, 83(6):2411–2451.

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

Ai, C. and Chen, X. (2012). The Semiparametric Efficiency Bound for Models of Sequential Moment Restrictions Containing Unknown Functions. *Journal of Econometrics*, 170(2):442–457.

Argañaraz, F. and Escanciano, J. C. (2023). On the Existence and Information of Orthogonal Moments For Inference. *arXiv preprint arXiv:2303.11418*.

References II

Athey, S., Imbens, G. W., and Wager, S. (2018). Approximate Residual Balancing: Debiased Inference of Average Treatment Effects in High Dimensions. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 80(4):597–623.

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

Belloni, A., Chen, D., Chernozhukov, V., and Hansen, C. (2012). Sparse Models and Methods for Optimal Instruments with an Application to Eminent Domain. *Econometrica*, 80(6):2369–2429.

Belloni, A., Chernozhukov, V., Fernández-Val, I., and Hansen, C. (2017). Program Evaluation and Causal Inference with High-Dimensional Data. *Econometrica*, 85(1):233–298.

References III

Bravo, F., Escanciano, J. C., and Keilegom, I. V. (2020). Two-Step Semiparametric Empirical Likelihood Inference. *The Annals of Statistics*, 48(1):1 – 26.

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

Chamberlain, G. (1987). Asymptotic Efficiency in Estimation with Conditional Moment Restrictions. *Journal of econometrics*, 34(3):305–334.

Chamberlain, G. (1992a). Comment: Sequential moment restrictions in panel data. *Journal of Business & Economic Statistics*, 10(1):20–26.

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

References IV

Chen, X. and Qiu, Y. J. J. (2016). Methods for Nonparametric and Semiparametric Regressions with Endogeneity: A Gentle Guide. *Annual Review of Economics*, 8:259–290.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W., and Robins, J. (2018). Double/debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*, 21:C1–C68.

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. M. (2022a). Locally Robust Semiparametric Estimation. *Econometrica*, 90(4):1501–1535.

Chernozhukov, V., Newey, W., and Robins, J. (2022b). De-Biased Machine Learning of Global and Local Parameters Using Regularized Riesz Representers. *arXiv preprint arXiv:1802.08667*.

References V

Chernozhukov, V., Newey, W., and Singh, R. (2022c). De-Biased Machine Learning of Global and Local Parameters Using Regularized Riesz Representers. *arXiv preprint arXiv:1802.08667*.

Chernozhukov, V., Newey, W., Singh, R., and Syrgkanis, V. (2020). Adversarial Estimation of Riesz Representers. *arXiv preprint arXiv:2101.00009*.

Chernozhukov, V., Newey, W. K., Quintas-Martinez, V., and Syrgkanis, V. (2021). Automatic Debiased Machine Learning Via Neural Nets for Generalized Linear Regression. *arXiv preprint arXiv:2104.14737*.

Chernozhukov, V., Newey, W. K., and Singh, R. (2022d). Automatic Debiased Machine Learning of Causal and Structural Effects. *Econometrica*, 90(3):967–1027.

Farrell, M. H. (2015). Robust Inference On Average Treatment Effects with Possibly More Covariates Than Observations. *Journal of Econometrics*, 189(1):1–23.

References VI

Farrell, M. H., Liang, T., and Misra, S. (2021). Deep Learning for Individual Heterogeneity: An Automatic Inference Framework. *arXiv preprint 2010.14694*.

Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007). Pathwise Coordinate Optimization. *The Annals of Applied Statistics*, 1(2):302 – 332.

Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models Via Coordinate Descent. *Journal of Statistical Software*, 33(1):1.

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

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

References VII

Ichimura, H. and Newey, W. K. (2022). The Influence Function of Semiparametric Estimators. *Quantitative Economics*, 13(1):29–61.

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

Nekipelov, D., Semenova, V., and Syrgkanis, V. (2022). Regularised Orthogonal Machine Learning for Nonlinear Semiparametric Models. *The Econometrics Journal*, 25(1):233–255.

Newey, W. K. and Powell, J. L. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5):1565–1578.

Olley, G. S. and Pakes, A. (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica*, 64(6):1263–1297.

Sasaki, Y. and Ura, T. (2023). Estimation and Inference for Policy Relevant Treatment Effects. *Journal of Econometrics*, 234(2):394–450.

References VIII

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