

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

Facundo Argañaraz

Universidad Carlos III de Madrid

June 18, 2024

Toulouse School of Economics

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 \rightarrow Standard inference is NOT typically valid.
- 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

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

- 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 l_{it} , e.g., investment, at t that is linked to ω_{it} :

$$l_{it} = l_t(\omega_{it}, X_{it}).$$

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

$$\omega_{it} = \omega_t(l_{it}, X_{it}).$$

Example: Production Functions III

- Equation (1) becomes

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

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

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

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

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

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

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

Production Functions IV

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

$$\mathbb{E} [Y_1 - \eta_{01}(l_1, X_1) | l_1, X_1] = 0,$$

$$\mathbb{E} [Y_2 - F(X_2, \theta_{0p}) - \theta_{0\omega} (\eta_{01}(l_1, X_1) - F(X_1, \theta_{0p})) | \Omega_1] = 0,$$

$$\mathbb{E} [Y_2 - \eta_{02}(l_2, X_2) | l_2, X_2] = 0,$$

$$\mathbb{E} [Y_3 - F(X_3, \theta_{0p}) - \theta_{0\omega} (\eta_{02}(l_2, X_2) - F(X_2, \theta_{0p})) | \Omega_2] = 0.$$

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

Production Functions V

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

$$\mathbb{E}[Y_1 - \eta_{01}(I_1, X_1) | I_1, X_1] = 0, \quad (2)$$

$$\mathbb{E}[Y_2 - F(X_2, \theta_{0p}) - \theta_{0\omega}(\eta_{01}(I_1, X_1) - F(X_1, \theta_{0p})) | \Omega_1] = 0, \quad (3)$$

$$\mathbb{E}[Y_2 - \eta_{02}(I_2, X_2) | I_2, X_2] = 0, \quad (4)$$

$$\mathbb{E}[Y_3 - F(X_3, \theta_{0p}) - \theta_{0\omega}(\eta_{02}(I_2, X_2) - F(X_2, \theta_{0p})) | \Omega_2] = 0. \quad (5)$$

- 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):

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

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

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

Figure: Comparison of Non-Orthogonal and Orthogonal Estimators

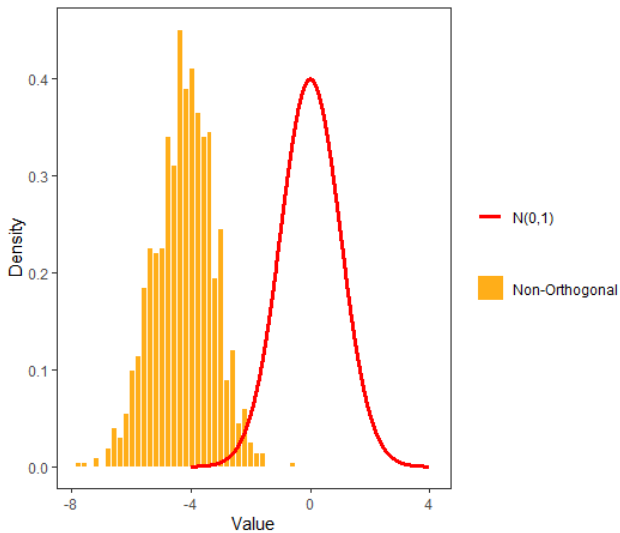
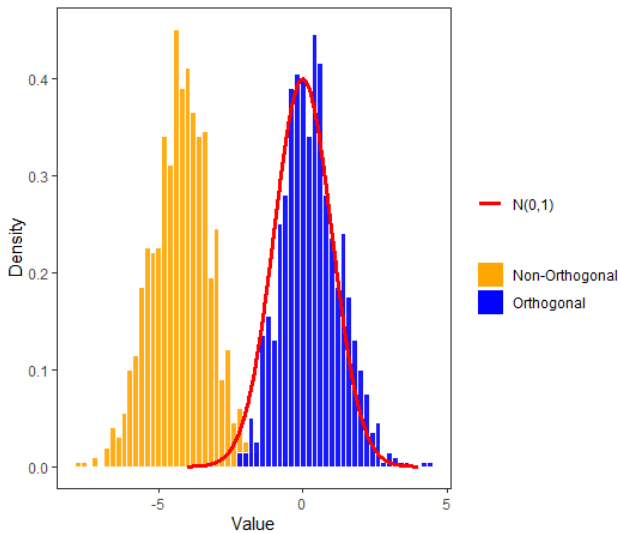


Figure: Comparison of Non-Orthogonal and Orthogonal Estimators



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 achieve 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 achieve 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 $\psi : \mathcal{W} \times \Theta \times \mathbf{B} \times L^2(Z) \mapsto \mathbb{R}$ satisfying the following two restrictions:

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

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

Related Literature

1 On debiased moments.

- Chernozhukov et al. (2022a); Athey et al. (2018); Belloni et al. (2012, 2017); Bravo et al. (2020); Farrell (2015); Nekipelov et al. (2022); Sasaki and Ura (2023); Ichimura and Newey (2022), **Chernozhukov et al. (2018)**, **Argañaraz and Escanciano (2023)**, ...

2 On models defined by CMRs.

- Chamberlain (1987, 1992b); Ai and Chen (2003); Chamberlain (1992a); Brown and Newey (1998), **Ai and Chen (2012)**; **Ackerberg et al. (2014)**, ...

3 On the automatic construction of debiased moments.

- Chernozhukov et al. (2022b,c, 2021, 2020); Farrell et al. (2021), **Chernozhukov et al. (2022d)**; **Bakhitov (2022)**, ...

Debiased Moments (continued)

- How can we construct ψ in our example?

Debiased Moments (continued)

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

$$\mathbb{E} [Y_1 - \eta_{01} (I_1, X_1) | I_1, X_1] = 0,$$

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

$$\mathbb{E} [Y_2 - \eta_{02} (I_2, X_2) | I_2, X_2] = 0,$$

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

Example (continued)

- We can obtain a debiased moment by means of

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

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

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

$$\begin{aligned}&\frac{d}{d\tau} \mathbb{E}[\psi(W, \theta_0, \eta_0 + \tau b, \kappa_0)] \\ &= \mathbb{E}[b_1(Z_1)(-\kappa_{01}(Z_1) - \theta_{0\omega}\kappa_{02}(Z_1)) + b_2(Z_2)(-\kappa_{02}(Z_2) - \theta_{0\omega}\kappa_{04}(Z_2))] \\ &= 0.\end{aligned}$$

HOW CAN WE GET κ_0 ?

How can we get κ_0 ? I

- Compute derivatives of each CMR:

$$\begin{aligned} \left[S_{\theta_0, \eta_0}^{(1)} b \right] (Z_1) &= -b_1(Z_1), & \left[S_{\theta_0, \eta_0}^{(2)} b \right] (Z_1) &= -\theta_{0\omega} b_1(Z_1), \\ \left[S_{\theta_0, \eta_0}^{(3)} b \right] (Z_2) &= -b_2(Z_2), & \left[S_{\theta_0, \eta_0}^{(4)} b \right] (Z_2) &= -\theta_{0\omega} b_2(Z_2). \end{aligned}$$

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

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

- For a valid κ_0 we need

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

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

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 $S_{\theta_0, \eta_0}^{(j)} S_{\theta_0, \eta_0}^*$ to approximate f^* , where S_{θ_0, η_0}^* is the adjoint of S_{θ_0, η_0} .

Estimation of OR-IVs II

- Let \mathcal{G} be some space of functions equipped with norm $\|\cdot\|_{\mathcal{G}}$ such that $\mathcal{G} \subseteq L^2(Z)$.
- In general, we are interested in solving

$$\min_{g \in \mathcal{G}} \sum_{j=1}^J \mathbb{E} \left[\left(f_j(Z_j) - S_{\theta_0, \eta_0}^{(j)} S_{\theta_0, \eta_0}^* g \right)^2 \right]. \quad (6)$$

- But...

- $S_{\theta_0, \eta_0}^{(j)} S_{\theta_0, \eta_0}^*$ is unknown \rightarrow Estimate it.
- Potentially, more than one solution exists \rightarrow Focus on the minimum norm solution g_0 .

Estimation of OR-IVs III

- I propose to estimate g_0 by means of

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

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

Estimation of OR-IVs IV

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

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

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

- In practical terms, we restrict the solution g_0 to be an element of \mathcal{G}_n in the sense that

$$\begin{aligned} \hat{S}_{\hat{\theta}, \hat{\eta}}^{(j)} \hat{S}_{\hat{\theta}, \hat{\eta}}^* g_0 &= \hat{S}_{\hat{\theta}, \hat{\eta}}^{(j)} \hat{S}_{\hat{\theta}, \hat{\eta}}^* \gamma_j(Z_j)' \beta_j \\ &\equiv \hat{M}_{j\ell} \beta, \end{aligned}$$

where the previous estimation is made with respect to fold ℓ .

Estimation of OR-IVs V

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

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

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

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

Estimation of OR-IVs - Recap

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

▶ Coordinate Descent Approach

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

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

Example (continued)

- The user has to provide

$f(Z) = (f_1(Z_1), f_2(Z_1), f_3(Z_2), f_4(Z_2)) \in L^2(Z)$ and basis $\gamma(Z_1)$ and $\gamma(Z_2)$.

- Regressors have the following expression

$$\left[\hat{\mathbf{M}}_{1\ell} \right]_{ik} = \gamma_k(Z_{1i}) + \hat{\theta}_{\omega\ell} \gamma_k(Z_{1i}), \quad (7)$$

$$\left[\hat{\mathbf{M}}_{2\ell} \right]_{ik} = \hat{\theta}_{\omega\ell} \left(\gamma_k(Z_{1i}) + \hat{\theta}_{\omega\ell} \gamma_k(Z_{1i}) \right), \quad (8)$$

$$\left[\hat{\mathbf{M}}_{3\ell} \right]_{ik} = \gamma_k(Z_{2i}) + \hat{\theta}_{\omega\ell} \gamma_k(Z_{2i}), \quad (9)$$

$$\left[\hat{\mathbf{M}}_{4\ell} \right]_{ik} = \hat{\theta}_{\omega\ell} \left(\gamma_k(Z_{2i}) + \hat{\theta}_{\omega\ell} \gamma_k(Z_{2i}) \right). \quad (10)$$

MORE IN THE PAPER

More in the paper I

1 A **general** setting ([▶ more details](#)):

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

where m_j can depend on θ_0 arbitrarily.

- Observe that if $Z_j = 1$, we have an unconditional moment.

Assumption

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

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

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

More in the paper II

Assumption

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

$$\left[S_{\theta_0, \eta_0}^{(j)} b \right] (Z_j) = \mathbb{E} [\nu_j (Y, \theta_0, \eta_0, b) | Z_j];$$

(ii) $\nu_j (Y, \theta_0, \eta_0, b) = b(X)' \tilde{\nu}_j (Y, \theta_0, \eta_0)$, for some d_η -vector of known (up to θ_0 and η_0) functions $\tilde{\nu}_j$.

(iii) For all $b \in \mathbf{B}$, $\mathbb{E} [b(X)' \tilde{\nu}_j (Y, \theta_0, \eta_0) | Z_j] \in L^2(Z_j)$.

Let $I_\ell^c = A_\ell + B_\ell + C_\ell$, under the previous assumptions, I show that

$$\hat{S}_{\hat{\theta}, \hat{\eta}}^{(j)} \hat{S}_{\hat{\theta}, \hat{\eta}}^* g = \hat{\mathbb{E}}_{C_\ell} \left[\left(\sum_{j'=1}^J \hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'} (Y, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}) g_{j'} (Z_j) | X \right] \right)' \tilde{\nu}_j (Y, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}) \middle| Z_j \right].$$

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. $\hat{\theta}_{A_\ell}$ and $\hat{\theta}_{B_\ell}$ are possibly non-LR estimators of θ_0 .

More in the paper III

2 Some ▶ regularity conditions are sufficient to show

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

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

■ Key conditions are:

More in the paper IV

Assumption

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

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

Assumption

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

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

More in the paper V

3 Introduce a GMM estimator $\hat{\theta}$ for θ_0 in a Two-Step setting.

- There are functions m_j 's that depend on η_0 only.
- Let $\hat{\eta}_\ell$ be an estimator of η_0 , using observations in I_ℓ^c . (► Endogeneity)
- Let

$$\hat{\psi}(\theta) = \frac{1}{n} \sum_{\ell=1}^L \sum_{i \in I_\ell} \psi(W_i, \theta, \hat{\eta}_\ell, \hat{\kappa}_\ell).$$

- Our proposed estimator $\hat{\theta}$ is defined as the solution to the GMM program

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \hat{\psi}(\theta)' \hat{\Lambda} \hat{\psi}(\theta), \quad (11)$$

More in the paper VI

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

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

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

$$\hat{V} = \left(\hat{\Upsilon}' \hat{\Lambda} \hat{\Upsilon} \right)^{-1} \hat{\Upsilon}' \hat{\Lambda} \hat{\Psi} \hat{\Lambda} \hat{\Upsilon} \left(\hat{\Upsilon}' \hat{\Lambda} \hat{\Upsilon} \right)^{-1}, \quad \hat{\Upsilon} = \frac{\partial}{\partial \theta} \hat{\psi}(\hat{\theta}). \quad (12)$$

- 4 Some regularity conditions are sufficient to show

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

- 5 $\hat{V} \xrightarrow{P} V.$

MONTÉ 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

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

- 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 $f(Z)$:

$$f_1(Z) = (K_{i1}, K_{i1}, K_{i2}, K_{i2})'$$

$$f_2(Z) = (l_{i1}, l_{i1}, l_{i2}, l_{i2})'$$

$$f_3(Z) = (K_{i1}, K_{i1}, l_{i2}, l_{i2})'$$

$$f_4(Z) = (K_{i1}, l_{i1}, l_{i2}, l_{i2})'$$

- 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)

Est.	$n = 500$						Random Forest
	Smaller λ_n	Larger λ_n	λ_n (BCCH)	Larger L	Larger r	Fourier Basis	
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

- This 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, I 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), Akerberg 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 $\ell = 1, \dots, L$, 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{v}_j , and thus can be evaluated.
- **Step 2:** Compute design matrix $\hat{\mathbf{M}}_{j\ell}$ such that its (i, l) -entry is

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

Algorithm to estimate OR-IVs II

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

$$\begin{aligned} [\hat{\mathbf{M}}_{j\ell}]_{il} &= \hat{\mathbb{E}}_{C_\ell} \left[\left(\hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'} \left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{jA_\ell} \right) \gamma_{j'k}^{low} \left(Z_{j'i} \right) \middle| X_i \right] \right)' \tilde{\nu}_j \left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{jB_\ell} \right) \middle| Z_{ji} \right], \\ \hat{\beta}_\ell &= \begin{pmatrix} \left(\sum_{j=1}^J \hat{\mathbf{M}}'_{j\ell} \hat{\mathbf{M}}_{j\ell} \right)^{-1} \left(\sum_{j=1}^J \hat{\mathbf{M}}'_{j\ell} \mathbf{f}_{j\ell} \right) \\ 0 \end{pmatrix} \end{aligned}$$

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

(a) Update normalization

$$\begin{aligned} \hat{\sigma}_{j'k\ell} &= \left[\frac{1}{n - n_\ell} \sum_{i \notin i_\ell} \left\{ \sum_{j=1}^J \hat{\mathbb{E}}_{C_\ell} \left[\left(\hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'} \left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{jA_\ell} \right) \gamma_{j'k}^{low} \left(Z_{j'i} \right) \middle| X_i \right] \right)' \tilde{\nu}_j \left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{jB_\ell} \right) \middle| Z_{ji} \right] \hat{\epsilon}_{j\ell} \right\}^2 \right]^{1/2} \\ \hat{\epsilon}_{j\ell} &= f_j(Z_{ji}) - \sum_{j'=1}^J \sum_{k=1}^K \hat{\beta}_{j'k\ell} \hat{\mathbb{E}}_{C_\ell} \left[\left(\hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'} \left(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{jA_\ell} \right) \gamma_{j'k}^{low} \left(Z_{j'i} \right) \middle| X_i \right] \right)' \tilde{\nu}_j \left(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{jB_\ell} \right) \middle| Z_{ji} \right]. \end{aligned}$$

Algorithm to estimate OR-IVs III

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

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

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

$$\begin{aligned} \hat{\kappa}_{j\ell}(Z_{ji}) &= f_j(Z_j) - \hat{f}_j^*(Z_j) \\ &= f_j(Z_{ji}) - \sum_{j'=1}^J \sum_{k=1}^{r_{j'}} \hat{\beta}_{j'k\ell} \hat{\mathbb{E}}_{C_\ell} \left[\left(\hat{\mathbb{E}}_{B_\ell} \left[\tilde{\nu}_{j'}(Y_i, \hat{\theta}_{A_\ell}, \hat{\eta}_{A_\ell}) \gamma_{j'k}(Z_{ji}) \mid X \right] \right)' \tilde{\nu}_j(Y_i, \hat{\theta}_{B_\ell}, \hat{\eta}_{B_\ell}) \mid Z_{ji} \right]. \end{aligned} \quad (13)$$

Coordinate Descent Approach I

- Step 4 of the iterative algorithm above requires to solve

$$\min_{\beta \in \mathbb{R}^r} \sum_{j=1}^J \frac{1}{n - n_\ell} \left(\mathbf{f}_{j\ell} - \hat{\mathbf{M}}_{j\ell} \beta \right)' \left(\mathbf{f}_{j\ell} - \hat{\mathbf{M}}_{j\ell} \beta \right) + 2\lambda_n \left\| \hat{\mathbf{D}}_\ell \beta \right\|_1, \quad (14)$$

- where $\hat{\mathbf{D}}_\ell$ is a diagonal matrix with elements $\hat{D}_{jkl} \equiv \hat{D}_{l\ell}$ along the main diagonal, with $l = 1, \dots, r$.
- 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 (14), 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):

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

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

Coordinate Descent Approach III

2 **Step 2:** Update coordinate β_l :

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

► Back

General Setting I

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

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

- m_j might depend on θ_0 arbitrarily.
- There exists a unique $(\theta_0, \eta_0) \in \Theta \times \mathbf{B}$ such that (50) holds.
- Let $\kappa = (\kappa_1, \dots, \kappa_J)$, where $\kappa_j \equiv \kappa_j(Z_j)$, and $\kappa_j \in L^2(Z_j)$.

General Setting II

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

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

Assumption

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

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

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

General Setting III

- Remark that (5) defines a linear operator $S_{\theta_0, \eta_0}^{(j)} : \mathbf{B} \mapsto L^2(Z_j)$. In addition, let us define

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

- $S_{\theta_0, \eta_0} : \mathbf{B} \mapsto L^2(Z)$ 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

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

- A key object:

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

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

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

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

Asymptotic results of OR-IVs I

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

$$\hat{F}_{j\ell} = \frac{1}{n - n_\ell} \sum_{i \notin I_\ell} f_j(Z_{ji}) \hat{\mathbf{M}}_{j\ell}(Z_{ji}), \quad F_j = \mathbb{E}[f_j(Z_j) \mathbf{M}_j(Z_j)],$$
$$\hat{G}_{j\ell} = \frac{1}{n - n_\ell} \sum_{i \notin I_\ell} \hat{\mathbf{M}}_{j\ell}(Z_{ji}) \hat{\mathbf{M}}_{j\ell}(Z_{ji})', \quad G_j = \mathbb{E}[\mathbf{M}_j(Z_j) \mathbf{M}_j(Z_j)'].$$

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

$$\hat{\beta}_\ell = \arg \min_{\beta \in \mathbb{R}^r} \sum_{j=1}^J \left(-2\hat{F}_{j\ell}'\beta - \beta' \hat{G}_{j\ell}\beta \right) + 2\lambda_n \|\beta\|_1. \quad (15)$$

Asymptotic results of OR-IVs II

Assumption

There are constants c_1, \dots, c_J such that with probability approaching one

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

Assumption

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

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

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

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

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

Asymptotic results of OR-IVs IV

Assumption

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

Assumption

Let

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

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

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 η_0

- I 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

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

- Consider the projected mean square norm:

$$\begin{aligned} \|T_j(\eta_j - \eta_{0j})\|_2 &= \sqrt{\mathbb{E}[\mathbb{E}[\eta_j(X) - \eta_{0j}(X) | Z_j]^2]}, \\ \|T(\eta - \eta_0)\|_{L^2(Z)} &\equiv \sqrt{\sum_{j=1}^J \|T_j(\eta_j - \eta_{0j})\|_2^2}. \end{aligned}$$

► Back

Asymptotic Results of D-CMRs I

Assumption

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

- i) $\int |m_j(y, \theta_0, \hat{\eta}_{j\ell}) - m_j(y, \theta_0, \eta_{0j})|^2 F_0(dw) \xrightarrow{P} 0$,
- ii) $\int |m_j(y, \theta_0, \hat{\eta}_{j\ell}) - m_j(y, \theta_0, \eta_{0j})|^2 \|\kappa_{0j}(z_j)\|^2 F_0(dw) \xrightarrow{P} 0$,
- iii) $\int |m_j(y, \theta_0, \eta_{0j})|^2 \|\hat{\kappa}_{j\ell}(z_j) - \kappa_{0j}(z_j)\|^2 \xrightarrow{P} 0$.

■ Let us define

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

Asymptotic Results of D-CMRs II

Assumption

There are constants c_1, \dots, c_J such that with probability approaching one

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

Assumption

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

Asymptotic Results of D-CMRs III

Assumption

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

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

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

$$i) \int \left\| \hat{\Delta}_\ell(w) \right\|^2 F_0(dw) \xrightarrow{P} 0, \quad \text{and} \quad ii) \sqrt{n} \int \hat{\Delta}_\ell(w) F_0(dw) \xrightarrow{P} 0. \quad (16)$$

Asymptotic Results of D-CMRs IV

- Let

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

Assumption

$\overline{\psi}(\theta, \eta, \kappa)$ 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

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

Asymptotic Results of D-CMRs V

- All the previous conditions are sufficient to show

$$\sqrt{n}\hat{\psi}(\theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi(W_i, \theta_0, \eta_0, \kappa_0) + o_p(1). \quad (17)$$

- The result in (17) is essential for obtaining asymptotic normality of $\hat{\theta}$.
- Interestingly, cross-fitting enables to show (17) 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

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

Asymptotic Results of D-CMRs VI

- We need conditions for convergence of the Jacobian:

$\frac{\partial}{\partial \theta} \hat{\psi}(\bar{\theta}) \xrightarrow{P} \Upsilon = \mathbb{E} \left[\frac{\partial}{\partial \theta} \psi(W, \theta_0, \eta_0, \kappa_0) \right]$ 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 κ_j) small enough, $\psi(W, \theta, \eta, \kappa)$ is differentiable in θ on \mathcal{N} with probability approaching one and there is a C and $d(W, \eta, \kappa)$ 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

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

- iii) For each q and k , $\int \left| \frac{\partial \psi_q(w, \theta_0, \hat{\eta}_\ell, \hat{\kappa}_\ell)}{\partial \theta_k} - \frac{\partial \psi_q(w, \theta_0, \eta_0, \kappa_0)}{\partial \theta_k} \right| F_0(dw) \xrightarrow{P} 0$.

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 18 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) $\mathbb{E} [\psi(W, \theta, \eta_0, \kappa_0)] = 0$ if and only if $\theta = \theta_0$; iii) Θ is compact; iv) $\int \left\| m_j(y, \theta, \hat{\eta}_{j\ell}) \hat{\kappa}_{j\ell}(\mathbf{z}_j) - m_j(y, \theta, \eta_{0j}) \kappa_{0j}(\mathbf{z}_j) \right\| F_0(dw) \xrightarrow{P} 0$ and $\mathbb{E} \left[\left\| m_j(Y, \theta, \eta_0) \kappa_{0j}(\mathbf{Z}_j) \right\|^2 \right] < \infty$ for all $\theta \in \Theta$; v) There is a $C > 0$ and $d(W, \eta, \kappa)$ such that for each $\|T(\eta - \eta_0)\|_{L^2(Z)} \|\kappa - \kappa_0\|_{L^2(Z)}$ small enough and all $\tilde{\theta}, \theta \in \Theta$,

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

Then, $\hat{\theta} \xrightarrow{P} \theta$.

Additional Monte Carlo Details I

- In our Monte Carlo experiments, I 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 I also consider a scenario where $L = 6$.
- 4 In a different experiment, I specify a larger number of coefficients such that $r = 25$.
- 5 Additionally, I 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 $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_k, \hat{\theta}_\omega)'$, I use GMM based on four debiased moments.
- These can be written as

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

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

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

Additional Monte Carlo Details III

- which implies that

$$\theta_{01} + \omega_t (I_t, K_t) = \eta_{0t} (Z_t) - \theta_{0k} K_t. \quad (18)$$

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

$$\omega_t = \theta_{0\omega} \omega_{t-1} + \epsilon_t^\omega, \quad \mathbb{E}[\epsilon_t^\omega | \omega_{t-1}] = 0. \quad (19)$$

- Plugging (18) into (19) and re-arranging terms yields

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

- 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 the Monte Carlo experiments, I directly consider $\tilde{c} = \theta_{01}$.
- 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, 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 Representer. *arXiv preprint arXiv:1802.08667*.
- Chernozhukov, V., Newey, W., Singh, R., and Syrgkanis, V. (2020). Adversarial Estimation of Riesz Representer. *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.