Automatic Orthogonal Moments for Production Functions Estimation

Facundo Argañaraz

Advisor: Juan Carlos Escanciano

Ph.D. Workshop

November 2023

Introduction

- Models defined by conditional moment restrictions (CMRs) are ubiquitous in economics.
  - Regressions, quantile models, missing data, simultaneous equations, rational expectations, and production functions, among many others (Chen and Qiu, 2016).
  - Despite their popularity, there does not exists a *general* and *automatic* approach that can be used to construct Locally Robust (LR)/Orthogonal/Debiased Moments (Chernozhukov et al., 2022; Neyman, 1959) for CMRs.

This paper

1. We introduce an **automatic** method to construct **debiased** moments in **general** semiparametric models defined by several CMRs, with possibly different conditioning variables and endogenous regressors.
2. We specialize our results in a fundamental model in economics: **production functions** at the firm level.

Why production functions?

- Production functions have been at the core of economics since the early 1800's (Chambers, 1988).
- Learning production functions (and productivity measures) is essential for answering several relevant questions for designing policies.
  - Trade liberalization, exporting, foreign ownership, competition, investment climate, learning by doing, to name a few (Ackerberg et al., 2007, 2015, ACF hereafter).

The general model

- $W_i = (Y_i, Z_i)$, $i = 1, \cdots, n$, is iid.
- Y is a random vector of endogenous variables taking values in $\mathcal{Y} \subseteq \mathbb{R}^{d_Y}$, and Z is random vector of exogenous variables taking values in $\mathcal{Z} \subseteq \mathbb{R}^{d_Z}$.
- Let $\theta \in \Theta \subset \mathbb{R}^{d_\theta}$ denote a finite-dimensional parameter vector.
- Let $\eta \in \boldsymbol{B}$ be a d_η-vector of real-valued measurable functions of W.

- There is a vector of residual functions $m_j : \mathcal{Y} \times \Theta \times \boldsymbol{B} \mapsto \mathbb{R}$ of Y, θ, and η, such that:

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

- where $\mathbb{E}[\cdot]$ is expectation under the distribution of Y given $Z^{(j)}$,
- μ_j is probability measure of $Z^{(j)}$,
- Z denotes the union of different random elements of $\left(Z^{(1)}, \cdots, Z^{(J)}\right)$
- m_j is known up to the parameters (θ_0, η_0).

- We are interested in settings that can be naturally labeled as semiparametric two-step.
- Specifically, when we write (1), we actually mean

$$\mathbb{E}\left[m_j\left(Y, \eta_{0j}\right) \middle| Z^{(j)}\right] = 0, \quad \mu_j - a.s., \quad j = 1, 2, \cdots, J_\eta, \tag{2}$$

$$\mathbb{E}\left[m_j\left(Y, \theta_0, \eta_0\right) \middle| Z^{(j)}\right] = 0, \quad \mu_j{-}a.s., \quad j = J_\eta{+}1, J_\eta{+}2, \cdots, J, \tag{3}$$

where $\eta_0 = \left(\eta_{0,1}, \cdots, \eta_{0,J_\eta}\right)$.

- We are not imposing anything regarding how the conditioning variables relate.
- Hereafter, we assume that there exists a unique $(\theta_0, \eta_0) \in \Theta \times \boldsymbol{B}$ such that (1) holds.

Learning Production Functions

- We observe a panel of n firms across T periods, where i and t index firms and periods, respectively. Note $W_i \equiv (W_{i1}, \cdots, W_{iT})$.
- Variables in the model:
  - Y_{1it}: "value-added" output;
  - X_{it}: inputs;
  - M_{it}: intermediate inputs.
- Output is determined by

$$Y_{1it} = p\left(X_{it}, \theta_{01}\right) + \omega_{it} + \epsilon_{it}, \tag{4}$$

- where ω_{it} is firm i's productivity shock (anticipated productivity) in period t, which is allowed to be correlated with inputs X_{it}, and ϵ_{it} is noise in output.

- Estimating production functions presents econometric challenges.
- Potential endogeneity between inputs and productivity (Marschak and Andrews, 1944).
- Several approaches have been proposed to solve this issue.
- Perhaps the most popular one in applied work is the proxy variable approach (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg et al., 2015; Wooldridge, 2009).
- ACF have highlighted identification issues in LP's model, which also carries over OP's setting.

- The assumptions of the model allows us to write

$$\omega_{it} = \mathcal{M}^{-1}\left(X_{it}, M_{it}\right). \tag{5}$$

- Then, we naturally obtain

$$\mathbb{E}\left[\left. Y_{1t} - \eta_0\left(X_t, M_t\right)\right| \Omega_t\right] = 0, \;\; a.s., \tag{FS}$$

$$\begin{aligned}\mathbb{E}[Y_{1t} - p\left(X_t, \theta_{01}\right) - h\left(\eta_0\left(X_{t-1}, M_{t-1}\right) - p\left(X_{t-1}, \theta_{01}\right), \theta_{02}\right) |\Omega_{t-1}] \\ = 0, \;\; a.s.,\end{aligned} \tag{SS}$$

- where where Ω_t is the firm's information set at t, and $\mathbb{E}\left[\omega_t | \Omega_{t-1}\right] = h(\omega_{t-1}, \theta_{02})$.
- Note $\theta_0 = (\theta_{01}, \theta_{02})$.

Debiased moments for CMRs

- Recall that $\eta_0 \in \boldsymbol{B}$. Let $\boldsymbol{B}$ be some Hilbert space with norm $||\cdot||_{\boldsymbol{B}}$ of measurable d_η-functions of W and inner product $\langle \cdot, \cdot \rangle_{\boldsymbol{B}}$.
- We assume that

$$g_j\left(Z^{(j)}, \theta_0, \eta\right) = \mathbb{E}\left[m_j\left(Y, \theta_0, \eta\right) \middle| Z^{(j)} \right] \tag{6}$$

- is "smooth" (at η_0) when viewed as a map from $\boldsymbol{B}$ into $L^2\left(Z^{(j)}\right)$.
- We assume that $g_j\left(Z^{(j)}, \theta_0, \cdot\right) : \boldsymbol{B} \mapsto L^2\left(Z^{(j)}\right)$ is Fréchet differentiable, where the derivative is given by

$$\begin{aligned} \left[\nabla g_j\left(Z^{(j)}, \theta_0, \eta_0\right)\right](b) &\equiv \frac{d}{d\tau} g_j\left(Z^{(j)}, \theta_0, \eta_0 + \tau b\right) \\ &= \left[S^{(j)}_{\theta_0, \eta_0} b\right]\left(Z^{(j)}\right), \end{aligned} \tag{7}$$

- for some $b \in \boldsymbol{B}$, where $\frac{d}{d\tau}$ denotes derivative from the right.

- We introduce the operator S_{θ_0,η_0} that is defined as follows

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

- Under suitable assumptions, results in Argañaraz and Escanciano (2023) implies that in our setting, LR moments are of the form

$$\psi\left(W, \theta_0, \eta_0, \kappa_0\right) = \sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0\right) \kappa_{0j}\left(Z^{(j)}\right), \tag{8}$$

- where $\kappa_0 = (\kappa_{01}, \cdots, \kappa_{0J}) \in \overline{\mathcal{R}}(S_{\theta_0,\eta_0})^{\perp}$. We denote them Orthogonal Instruments (O-IVs).
- How can we get those O-IVs?

Debiased moments for production functions

- Let us assume that $(X_t, M_t) \subset \Omega_t$. Let $C_t = (X_t, M_t, X_{t-1}, M_{t-1})$.
- A simple derivation indicates that $S_{\theta_0,\eta_0} : \boldsymbol{B} \mapsto L^2(C_t) \times L^2(C_t)$ is as follows

$$S_{\theta_0,\eta_0} b = (-b(C_t), -h_\omega(\eta_{0,t-1}(X_{t-1}, M_{t-1}) - p(X_{t-1}, \theta_{01}), \theta_{02}) b(C_t)) \tag{9}$$

- where h_ω is the derivative of h with respect to ω_{t-1}.

Proposition

A LR moment for the model of production function estimation, introduced by Olley and Pakes (1996), is given by

$$\begin{aligned}\psi\left(W_t,\theta_0,\eta_0,\kappa_0\right) = {} & \left(Y_{1,t}-\eta_0\left(X_t,M_t\right)\right)\kappa_{01}\left(C_t\right) \\ & +\left(Y_{1t}-p\left(X_t,\theta_{01}\right)-\right. \\ & \left. h\left(\eta_0\left(X_{t-1},M_{t-1}\right)-p\left(X_{t-1},\theta_{01}\right),\theta_{02}\right)\right)\kappa_{02}\left(C_t\right),\end{aligned} \tag{10}$$

where $\kappa_0=\left(\kappa_{01}\left(C_t\right),\kappa_{02}\left(C_t\right)\right)\in\overline{\mathcal{R}}\left(S_{\theta_0,\eta_0}\right)^{\perp}$.

Estimation of the O-IVs

- Suppose that there exists a function ν_j such that the Fréchet derivative of $\mathbb{E}\left[\psi\left(W, \theta_0, \eta, \kappa_0\right)\right]$ in the direction b is

$$\begin{aligned}
\frac{d}{d\tau}\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0 + \tau b, \kappa_0\right)\right] &= \frac{d}{d\tau}\mathbb{E}\left[\sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0 + \tau b\right)\kappa_{0j}\left(Z^{(j)}\right)\right. \\
&= \sum_{j=1}^{J}\mathbb{E}\left[\nu_j\left(Y, \theta_0, \eta_0, b\right)\kappa_{0j}\left(Z^{(j)}\right)\right] \\
&= 0,
\end{aligned} \tag{11}$$

- for any $b \in \boldsymbol{B}$.
- Typically, ν_j will be obtained by direct calculation.

- Let us assume that for all $1 \leq j \leq J$, the estimator $\hat{\kappa}_j$ is of the form $\hat{\kappa}_j\left(Z^{(j)}\right) = \gamma_j\left(Z^{(j)}\right)' \hat{\beta}_j$, where $\gamma_j\left(Z^{(j)}\right)$ is a r_j-dimensional vector of basis functions that have been standardized to have mean zero and standard deviation 1.
- Let $D(W)$ be a $d_\eta \times q_1$ matrix of basis functions for deviations b.
- Let $d_s \equiv d_s(W)$ be the $s-$column of $D(W)$. Note that d_s belongs to $\boldsymbol{B}$.
- We can construct a sample analog of the derivative in (11) by replacing b for d_s and κ_{0j} for $\gamma\left(Z^{(j)}\right)' \beta_j$, leading to

$$\hat{\psi}_{\eta\ell}(d_s, \beta_\ell) = \frac{1}{n - n_\ell} \sum_{\ell' \neq \ell} \sum_{i \in I_{\ell'}} \sum_{j=1}^{J} \nu_j\left(Y, \tilde{\theta}, \hat{\eta}_{\ell\ell'}, d_s\right) \gamma_j\left(Z_i^{(j)}\right)' \beta_{j\ell},$$

$$s = 1, \cdots, q_1. \tag{12}$$

PGMM

- Implement the penalized GMM (PGMM) framework, following Caner and Kock (2019) and Bakhitov (2022).
- Let $\hat{\psi}_{\eta\ell}\left(\hat{\beta}_\ell\right) = \left(\hat{\psi}_{\eta\ell}\left(d_1, \hat{\beta}_\ell\right), \cdots, \hat{\psi}_{\eta\ell}\left(d_{q_1}, \hat{\beta}_\ell\right)\right)'$, where each $\hat{\psi}_{\eta\ell}\left(d_s, \hat{\beta}_\ell\right)$ is given in (12).
- The solution to the PGMM program is defined as

$$\hat{\beta}_\ell := \underset{\beta \in \mathbb{R}^r}{\arg\min}\, \hat{\psi}_{\eta\ell}(\beta)' \hat{\Lambda}_{q_1} \hat{\psi}_{\eta\ell}(\beta) + 2\lambda_{1n} \|\beta\|_1 - 2\lambda_{2n} \|\beta\|_1,$$

- where $\hat{\Lambda}_{q_1} = \hat{\Lambda}/q_1$, $\hat{\Lambda}$ is a $q_1 \times q_1$ positive semi-definite matrix, λ_{1n} and λ_{2n} are tuning parameters that depend on n and should satisfy $\lambda_{1n} \geq \lambda_{2n}$, and $\|\cdot\|_1$ is the L_1-norm.

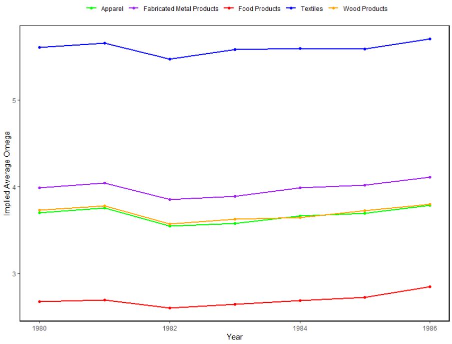
Data

- Data from Instituto Nacional de Estadistica de Chile.
- Information on all Chilean manufacturing plants with at least ten employees in the period 1979-1986.
- We focus on the five largest three-digit (ISIC codes) manufacturing industries in Chile: food products (311), textiles (321), apparel (322), wood products (331), and fabricated metal products (381).
- Plant variables are collected annually and they include revenues, investment, capital formation, different types of labor (blue and white collar), and measures of intermediate inputs (materials, services, electricity, and fuels).

Table 3: Empirical Results by 3-digit sector

Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
All ($n = 1{,}223$)				
$\hat{\theta}_k$	0.800	0.018	0.488	0.000
$\hat{\theta}_l$	0.200	0.007	0.457	0.001
$\hat{\theta}_2$	0.901	0.004	1.005	0.001
Food products ($n = 689$)				
$\hat{\theta}_k$	0.546	0.001	0.451	0.000
$\hat{\theta}_l$	0.519	0.001	0.619	0.000
$\hat{\theta}_2$	0.996	0.005	1.006	0.001
Textiles ($n = 158$)				
$\hat{\theta}_k$	0.098	0.297	0.272	0.000
$\hat{\theta}_l$	0.899	0.188	0.826	0.000
$\hat{\theta}_2$	0.823	0.192	1.003	0.001
Apparel ($n = 116$)				
$\hat{\theta}_k$	0.505	2.714	0.231	0.000
$\hat{\theta}_l$	0.513	1.829	1.034	0.000
$\hat{\theta}_2$	0.317	0.310	0.998	0.002
Wood Products ($n = 118$)				
$\hat{\theta}_k$	0.400	2.616	0.080	0.000
$\hat{\theta}_l$	0.600	1.487	1.243	0.000
$\hat{\theta}_2$	0.116	0.031	0.996	0.002
Fabricated Metal Products ($n = 142$)				
$\hat{\theta}_k$	0.401	1.569	0.274	0.000
$\hat{\theta}_l$	0.602	0.742	0.902	0.000
$\hat{\theta}_2$	0.551	0.258	1.002	0.001

Figure: Implied productivity



Final Rermarks

- We have introduced an automatic construction of debiased moments for general models defined by several conditioning moment restrictions, with possible different conditioning variables and endogenous regressors.
- We have leveraged the generality of our theory to derive orthogonal moments in the context of production functions estimation at the firm level, following the approach of ACF.

Appendix

Estimation of the parameter of interest

- To estimate the parameter of interest θ_0, we follow Chernozhukov et al. (2018) and use cross-fitting.
- Let I_ℓ, $\ell = 1, \cdots, L$, be a random partition of the observation index set $\{1, \cdots, n\}$ into L distinct subsets of about the same size.
- Let $\hat{\eta}_\ell$ and $\hat{\kappa}_\ell$ be given estimators of η_0 and κ_0, respectively, based on observations that are not in I_ℓ.
- The *CMRs Debiased Machine Learning Estimator* (CMRs-D-ML) is

$$\hat{\theta} := \arg\min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{\ell=1}^{L} \sum_{i \in I_\ell} \psi\left(W_i, \theta, \hat{\eta}_\ell, \hat{\kappa}_\ell\right) \right)^2 . \tag{13}$$

Monte Carlo

- To produce output, firms employ three inputs, namely, L_1, L_2, and K, with a Cobb-Douglas technology.
- The parameters θ_{0l_1}, θ_{0l_2}, and θ_{0k} are the corresponding input elasticities.
- The level of capital is generated by the perpetual inventory method.
- Productivity ω_t follows an AR(1) process with persistence parameter given by θ_{02}.
- There exists an intermediate input that is assumed to follow:

$$M_{it} = \exp\left\{-0.5 u_{it} u_{it}' + \omega_{it}\right\},$$

- where $u_{it} = (L_{it}', K_{it})'$.
- We consider the data from the steady state distribution implied by the model.

Table 1: Monte Carlo Results

True Parameters: $\theta_{0k} = 0.4$, $\theta_{0l_1} = 0.3$, $\theta_{0l_2} = 0.3$, $\theta_{02} = 0.7$						
			$n = 100$			
Est.	Coef. (CMRs-D-ML)	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.428	0.028	0.018	0.281	-0.119	1.109
$\hat{\theta}_{l_1}$	0.311	0.011	0.024	0.335	0.035	0.539
$\hat{\theta}_{l_2}$	0.310	0.010	0.024	0.322	0.022	0.516
$\hat{\theta}_2$	0.666	-0.034	0.053	0.905	0.205	0.101
			$n = 500$			
Est.	Coef. (CMRs-D-ML)	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.424	0.024	0.012	0.360	-0.040	0.542
$\hat{\theta}_{l_1}$	0.310	0.010	0.013	0.289	-0.011	0.277
$\hat{\theta}_{l_2}$	0.306	0.006	0.016	0.316	0.016	0.284
$\hat{\theta}_2$	0.676	-0.024	0.025	0.863	0.163	0.097
			$n = 1,000$			
Est.	Coef. (CMRs-D-ML)	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.422	0.022	0.011	0.348	-0.052	0.385
$\hat{\theta}_{l_1}$	0.309	0.009	0.013	0.303	0.003	0.178
$\hat{\theta}_{l_2}$	0.305	0.005	0.014	0.303	0.003	0.164
$\hat{\theta}_2$	0.680	-0.020	0.019	0.833	0.133	0.097

References I

Ackerberg, D., Benkard, C. L., Berry, S., and Pakes, A. (2007). Econometric Tools for Analyzing Market Outcomes. *Handbook of econometrics*, 6:4171–4276.

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

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

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

Caner, M. and Kock, A. B. (2019). High Dimensional Linear GMM. *arXiv preprint arXiv:1811.08779*.

References II

Chambers, R. G. (1988). *Applied Production Analysis: A Dual Approach*. Cambridge University Press.

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. (2022). Locally Robust Semiparametric Estimation. *Econometrica*, 90(4):1501–1535.

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

References III

Marschak, J. and Andrews, W. H. (1944). Random Simultaneous Equations and the Theory of Production. *Econometrica, Journal of the Econometric Society*, pages 143–205.

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

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

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