Automatic Orthogonal Moments for Production Functions Estimation

Facundo Argañaraz

Advisor: Juan Carlos Escanciano

Applied Reading Group

UC3M

December 2023

What the title means?

- Automatic?
- Orthogonal Moments?
- Production functions?

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 to answer 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).

Learning Production Functions

- We observe a panel of n firms across T periods, where i and t index firms and periods, respectively.
- Output is determined by

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

- where p is known up to θ_{01}, 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).

The Proxy Variable Approach

- Let Ω_t be the firm's information set at t.
- We assume that productivity follows a First-Order Markov Process, i.e.

$$\mathbb{E}\left[\omega_t | \Omega_{t-1}\right] = h(\omega_{t-1}, \theta_{02}),$$

- where h is known up to θ_{02}.
- In addition, from the firm's dynamic optimization problem, there exists a link

$$M_t = \mathcal{M}\left(X_t, \omega_t\right).$$

- The previous relationship might be hard to fully characterize and known in practice, unless more restrictive assumptions are imposed.
- Under regularity assumptions, we can write

$$\omega_{it} = \mathcal{M}^{-1}\left(X_{it}, M_{it}\right). \tag{2}$$

- The above implies that we have

$$\begin{aligned} Y_{1it} &= p\left(X_{it}, \theta_{01}\right) + \mathcal{M}^{-1}\left(X_{it}, M_{it}\right) + \epsilon_{it} \\ &= \eta_0\left(X_{it}, M_{it}\right) + \epsilon_{it}. \end{aligned} \tag{3}$$

- We cannot identify θ_{01}, but...

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

- In addition,

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

- Estimate η_0 using (FS) and plug it into (SS).

Machine Learning

- The standard practice in applied work uses a low dimensional polynomial to estimate η_0.
- The above might impose heavy parametric restrictions on η_0, and thus in our model. There is no reason to believe that these are correct.
- More conveniently, we shall use **machine learning** tools to estimate η_0 as these are powerful (i.e., flexible) techniques to learn objects such as η_0.
- However, this possibility brings up new issues...

First-stage bias

- Machine learning tools will typically produce biased estimates.
- This is fine if we want to predict, but we want to conduct a causal analysis on θ_0.
- The problem is that first-stage bias is transmitted to the second stage, which induces a bias on $\hat{\theta}_0$, invalidating standard inference.

Orthogonal Moments

- To get rid of the bias appearing in the second stage, we use Orthogonal Moments.
- An orthogonal/debiased/locally robust moment is less affected by this bias.
- In general, suppose that θ_0 "solves"

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

- The moment is orthogonal to η_0 when

$$\frac{d}{d\tau}\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0 + \tau b\right)\right] = 0.$$

- Notice that ψ appears in an unconditional moment, and we might need to derive ψ.

A special setting

- Production functions are a special model.
- Conditional Moment Restrictions (CMRs).
  - Select, in an ad-hoc way, instruments to have an unconditional moment? But, we have a (VERY) large number of potential instruments.
- This is a non-trivial model. It might be impossible to find a closed-form expression for ψ.
- So... what then?

Automatic construction

- Instead of finding a closed form expression for ψ, **estimate** it from the data, in a flexible way.
- Our method automatically selects, among the large set of possible instruments, those that would derive in an orthogonal (unconditional) moment.
- Use a "special" GMM approach.
    - As such, the estimation is easy to implement.

How we do it?

- Suppose we have that our θ_0 "solves":

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

- It turns out that, in our setting, orthogonal moments are of the form

$$\boxed{\psi\left(W, \theta_0, \eta_0, \kappa_0\right) = \sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0\right) \kappa_{0j}\left(Z^{(j)}\right)} \tag{4}$$

- for very special functions $\kappa_0 = \left(\kappa_{01}, \cdots, \kappa_{0J}\right)$, which we denote Orthogonal Instruments (O-IVs).

So?

- We find a moment restriction for κ_0.
- We "assume" that $\kappa_{0j} = \gamma_j \left(Z^{(j)}\right)' \beta_j$.
- Estimate each β_j with GMM.

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

$$\hat{\theta} := \arg\min_{\theta \in \Theta} \left(\frac{1}{n} \sum_{\ell=1}^{L} \sum_{i \in I_\ell} \psi\left(W_i, \theta, \hat{\eta}_\ell, \hat{\kappa}_\ell\right) \right)^2. \tag{5}$$

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.

Specifications matter after all?

Table 5: Sensitivity Results

True Parameters: $\theta_{0k} = 0.4$, $\theta_{0l_1} = 0.3$, $\theta_{0l_2} = 0.3$, $\theta_{02} = 0.7$				
Degree	$\hat{\theta}_k$	$\hat{\theta}_{l_1}$	$\hat{\theta}_{l_2}$	$\hat{\theta}_2$
1	0.594	0.229	0.244	0.926
	[-0.417, 1.605]	[-0.321, 0.779]	[-0.277, 0.766]	[0.830, 1.021]
2	0.359	0.330	0.351	0.843
	[-0.600, 1.317]	[-0.025, 0.685]	[-0.074, 0.777]	[0.648, 1.037]
3	0.320	0.313	0.325	0.857
	[-0.684, 1.324]	[-0.160, 0.786]	[-0.127, 0.777]	[0.658, 1.056]
4	0.449	0.279	0.283	0.863
	[-0.539, 1.438]	[-0.231, 0.790]	[-0.192, 0.758]	[0.679, 1.047]
5	0.456	0.270	0.286	0.863
	[-0.531, 1.443]	[-0.244, 0.784]	[-0.216, 0.788]	[0.680, 1.045]

Bias is a problem after all?

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.	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.029	0.016	-0.119	1.109
$\hat{\theta}_{l_1}$	0.010	0.023	0.035	0.539
$\hat{\theta}_{l_2}$	0.011	0.022	0.022	0.516
$\hat{\theta}_2$	-0.039	0.048	0.205	0.101
		$n = 500$		
Est.	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.025	0.013	-0.040	0.542
$\hat{\theta}_{l_1}$	0.010	0.015	-0.011	0.277
$\hat{\theta}_{l_2}$	0.010	0.016	0.016	0.284
$\hat{\theta}_2$	-0.030	0.031	0.163	0.097
		$n = 1,000$		
Est.	Bias (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Bias (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.023	0.012	-0.052	0.385
$\hat{\theta}_{l_1}$	0.008	0.014	0.003	0.178
$\hat{\theta}_{l_2}$	0.010	0.014	0.003	0.164
$\hat{\theta}_2$	-0.027	0.025	0.133	0.097

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 2: Some descriptive statistics on Chilean manufacturing plants

Industry (ISIC)	Plants	Value Added	Capital	Labor	Intermediate Input
Food products (311)	689	9.684	9.106	3.571	10.757
		(1.587)	(1.865)	(0.840)	(1.400)
Textiles (321)	158	10.439	10.087	3.684	10.845
		(1.587)	(1.914)	(1.180)	(1.413)
Apparel (322)	116	9.887	8.881	3.412	10.475
		(1.371)	(1.428)	(0.850)	(1.249)
Wood Products (331)	118	9.758	9.468	3.525	10.437
		(1.404)	(1.672)	(0.841)	(1.173)
Fabricated Metal Products (381)	142	10.593	9.850	3.752	10.731
		(1.518)	(1.890)	(0.931)	(1.438)

Table 3: Empirical Results by 3-digit sector

All (n = 1,223)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.809	0.084	0.488	0.121
$\hat{\theta}_l$	0.206	0.036	0.457	0.336
$\hat{\theta}_2$	0.902	0.038	1.005	0.003
$\hat{\theta}_k + \hat{\theta}_l$	1.015	0.083	0.945	0.217
Food Products (n = 689)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.704	0.109	0.451	0.131
$\hat{\theta}_l$	0.322	0.043	0.619	0.364
$\hat{\theta}_2$	0.904	0.039	1.006	0.003
$\hat{\theta}_k + \hat{\theta}_l$	1.026	0.097	1.070	0.235
Textiles (n = 158)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.193	0.052	0.273	0.070
$\hat{\theta}_l$	0.802	0.024	0.821	0.124
$\hat{\theta}_2$	0.525	0.028	1.003	0.001
$\hat{\theta}_k + \hat{\theta}_l$	0.995	0.068	1.094	0.074
Apparel (n = 116)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.372	0.087	0.236	0.116
$\hat{\theta}_l$	0.593	0.053	1.027	0.380
$\hat{\theta}_2$	0.524	0.045	0.998	0.002
$\hat{\theta}_k + \hat{\theta}_l$	0.965	0.133	1.263	0.301

Table 4: Empirical Results by 3-digit sector (continued)

Wood Products (n = 118)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.231	0.165	0.081	0.199
$\hat{\theta}_l$	0.703	0.092	1.233	0.618
$\hat{\theta}_2$	0.920	0.197	0.996	0.004
$\hat{\theta}_k + \hat{\theta}_l$	0.934	0.171	1.314	0.433
Fabricated Metal Products (n = 142)				
Est.	Coef. (CMRs-D-ML)	Std. Err. (CMRs-D-ML)	Coef. (ACF)	Std. Err. (ACF)
$\hat{\theta}_k$	0.147	0.127	0.272	0.074
$\hat{\theta}_l$	0.953	0.064	0.907	0.201
$\hat{\theta}_2$	0.780	0.130	1.002	0.001
$\hat{\theta}_k + \hat{\theta}_l$	1.100	0.144	1.179	0.134

Figure 1: Implied Average Productivity (Omega) by 3-digit industry during 1980-1986

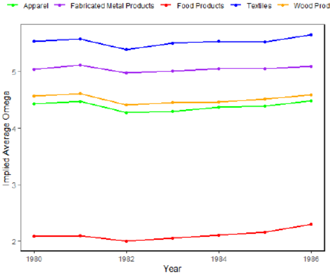
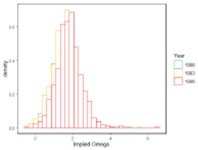
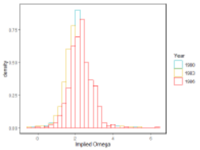


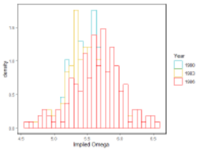
Figure 2: Distribution of Implied Omegas by Industry and Year



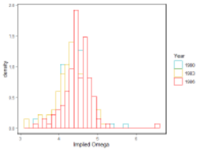
(a) All



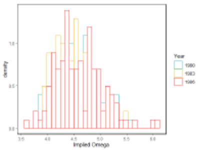
(b) Food Products



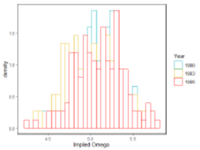
(c) Textiles



(d) Apparel



(e) Wood Products



(f) Fabricated Metal Products

Final Remarks

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.

Appendix

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:

$$\mathbb{E}\left[m_j\left(Y, \theta_0, \eta_0\right) \middle| Z^{(j)} \right] = 0, \quad \mu_j - a.s., \quad j = 1, 2, \cdots, J, \qquad (6)$$

- 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).

- Specifically, when we write (6), we actually mean

$$\mathbb{E}\left[\left. m_j\left(Y,\eta_{0j}\right)\right| Z^{(j)}\right]=0,\quad \mu_j-a.s.,\quad j=1,2,\cdots,J_\eta, \tag{7}$$

$$\mathbb{E}\left[\left. m_j\left(Y,\theta_0,\eta_0\right)\right| Z^{(j)}\right]=0,\quad \mu_j-a.s.,\quad j=J_\eta+1,J_\eta+2,\cdots,J, \tag{8}$$

  where $\eta_0=\left(\eta_{01},\cdots,\eta_{0J_\eta}\right)$.

- Hereafter, we assume that there exists a unique $(\theta_0,\eta_0)\in\Theta\times\boldsymbol{B}$ such that (7) and (8) hold.

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

$$g_j\left(Z^{(j)}, \theta_0, \eta\right) = \mathbb{E}\left[m_j\left(Y, \theta_0, \eta\right) \middle| Z^{(j)} \right] \tag{9}$$

- is "smooth" (at η_0).
- 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

$$\begin{aligned} \left[\nabla g_j\left(Z^{(j)}, \theta_0, \eta_0\right)\right](b) &\equiv \frac{d}{d\tau} g_j\left(Z^{(j)}, \theta_0, \eta_0 + \tau b\right) \\ &= \left[S^{(j)}_{\theta_0, \eta_0} b\right]\left(Z^{(j)}\right), \end{aligned} \tag{10}$$

- 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) imply that in our setting, LR moments are of the form

$$\boxed{\psi\left(W, \theta_0, \eta_0, \kappa_0\right) = \sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0\right) \kappa_{0j}\left(Z^{(j)}\right),} \tag{11}$$

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

$$\overline{\mathcal{R}}\left(S_{\theta_0,\eta_0}\right)^{\perp} = \left\{ f_1 \in \bigotimes_{j=1}^{J} L^2\left(Z^{(j)}\right) : \mathbb{E}\left[f_1' f_2\right] = 0, f_2 \in \overline{\mathcal{R}}\left(S_{\theta_0,\eta_0}\right) \right\}.$$

- How can we get those O-IVs?

Debiased moments for production functions

- Let us assume that $(X_t, M_t) \subset \Omega_t$.
- A simple derivation indicates that

  $S_{\theta_0,\eta_0} : \boldsymbol{B} \mapsto L^2(X_t, M_t) \times L^2(X_{t-1}, M_{t-1})$ is as follows

$$S_{\theta_0,\eta_0} b = (-b(X_t, M_t), -h_\omega(\eta_0(X_{t-1}, M_{t-1}) - p(X_{t-1}, \theta_{01}), \theta_{02}) \\ b(X_{t-1}, M_{t-1})), \tag{12}$$

- 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

$$\begin{aligned}\psi\left(W_t,\theta_0,\eta_0,\kappa_0\right) = {} & \left(Y_{1t}-\eta_0\left(X_t,M_t\right)\right)\kappa_{01}\left(X_t,M_t\right) \\ & + \left(Y_{1t}-p\left(X_t,\theta_{01}\right)-h\left(\eta_0\left(X_{t-1},M_{t-1}\right)-\right.\right. \\ & \left.\left. p\left(X_{t-1},\theta_{01}\right),\theta_{02}\right)\kappa_{02}\left(X_{t-1},M_{t-1}\right)\right),\end{aligned} \tag{13}$$

where $\kappa_0 = \left(\kappa_{01}\left(X_t,M_t\right),\kappa_{02}\left(X_{t-1},M_{t-1}\right)\right) \in \overline{\mathcal{R}}\left(S_{\theta_0,\eta_0}\right)^{\perp}$.

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

$$
\begin{aligned}
\frac{d}{d\tau}\mathbb{E}\left[\psi\left(W, \theta_0, \eta_0 + \tau b, \kappa_0\right)\right] &= \frac{d}{d\tau}\mathbb{E}\left[\sum_{j=1}^{J} m_j\left(Y, \theta_0, \eta_0 + \tau b\right)\kappa_{0j}\left(Z^{(j)}\right)\right. \\
&= \sum_{j=1}^{J}\mathbb{E}\left[\nu_j\left(Y, \theta_0, \eta_0, b\right)\kappa_{0j}\left(Z^{(j)}\right)\right] \\
&= 0,
\end{aligned}
\tag{14}
$$

- 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.
- 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 (14) by replacing b for d_s and κ_{0j} for $\gamma\left(Z^{(j)}\right)' \beta_j$, leading to

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

$$s = 1, \cdots, q_1. \tag{15}$$

PGMM

- Implement the penalized GMM (PGMM) framework, following Caner and Kock (2019) and Bakhitov (2022).
- Let us define, with some abuse of notation, the $q_1 \times r$ matrix $\hat{G}_\ell$ as follows:

$$\hat{G}_\ell := \begin{bmatrix} \frac{1}{n-n_\ell}\sum_i \nu_1\left(Y_i, \tilde{\theta}, \hat{\eta}_{\ell\ell'}, d_1\right)\gamma_1' & \cdots & \frac{1}{n-n_\ell}\sum_i \nu_J\left(Y_i, \tilde{\theta}, \hat{\eta}_{\ell\ell'}, d_1\right)\gamma_J' \\ & \cdots & \\ \frac{1}{n-n_\ell}\sum_i \nu_1\left(Y_i, \tilde{\theta}, \hat{\eta}_{\ell\ell'}, d_{q_1}\right)\gamma_1' & \cdots & \frac{1}{n-n_\ell}\sum_i \nu_J\left(Y_i, \tilde{\theta}, \hat{\eta}_{\ell\ell'}, d_{q_1}\right)\gamma_J' \end{bmatrix}$$

- The PGMM program is

$$\min_{\beta \in \mathbb{R}^r} \left(\hat{G}_\ell \beta\right)' \hat{\Lambda}_{q_1} \left(\hat{G}_\ell \beta\right) \quad \text{s.t. } ||\beta||_1 \leq c_1 \text{ and } ||\beta||_1 \geq c_2, \quad (16)$$

- where $\hat{\Lambda}_{q_1} = \hat{\Lambda}/q_1$, $\hat{\Lambda}$ is a $q_1 \times q_1$ positive semi-definite matrix, and c_1 and c_2 are positive constants.

- The solution can be written as

$$\hat{\beta}_\ell := \arg\min_{\beta \in \mathbb{R}^r} \left(\hat{G}_\ell \beta\right)' \hat{\Lambda}_{q_1} \left(\hat{G}_\ell \beta\right) + 2\lambda_{1n} \left|\left|\beta\right|\right|_1 - 2\lambda_{2n} \left|\left|\beta\right|\right|_1 .$$

- where λ_{1n} and λ_{2n} are tuning parameters that depend on n and should satisfy $\lambda_{1n} \geq \lambda_{2n}$.

What is PGMM doing?

- Coordinate Descent Algorithm.
- The solution is

$$\hat{\beta} = \underset{\beta \in \mathbb{R}^r}{\arg\min} \frac{1}{2} (G\beta)' \Lambda_{q_1} (G\beta) + \lambda_n ||\beta||_1 ,$$

- where we define $\lambda_n := \lambda_{1n} - \lambda_{2n}$.
- The derivative of the first term of the objective function with respect to β_j is

$$\begin{aligned} \frac{\partial}{\partial \beta_j} \left[\frac{1}{2} (G\beta)' \Lambda_{q_1} (G\beta) \right] &= e_j' G' \Lambda_{q_1} (G\beta - Ge_j\beta_j + Ge_j\beta_j) \\ &= A_j + B_j\beta_j. \end{aligned}$$

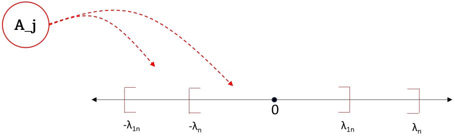
- The subgradient of the penalty term is

$$\frac{\partial}{\partial \beta_j} \lambda_n \left|\left|\beta\right|\right|_1 = \begin{cases} -\lambda_n & \text{if } \beta_j < 0 \\ [-\lambda_n, \lambda_n] & \text{if } \beta_j = 0 \\ \lambda_n & \text{if } \beta_j > 0 \end{cases}$$

- The coordinate solution can be computed as

$$\beta_j = \begin{cases} \frac{\lambda_n - A_j}{B_j} & \text{if } \lambda_n < A_j \\ 0 & \text{if } A_j \in [-\lambda_n, \lambda_n] \\ \frac{-\left(\lambda_n + A_j\right)}{B_j} & \text{if } -\lambda_n > A_j \end{cases}$$

- Our implementation does not prevent the trivial solution if tuning parameters are not properly chosen.
- There are two reasons why we still might end up with a trivial solution.
  1. When the rest of the coordinates tend to the trivial solution, A_j approaches zero, and β_j will be set to zero, and thus between $-\lambda_n$ and λ_n.
  2. If λ_{1n} is large enough. This is typical in any "Lasso" problem".
- To avoid the above from happening, the second tuning parameter, λ_{2n}, reduces λ_{1n}, making it more plausible to pick up a solution different from the trivial one.



- Thus, we are able to find a feasible solution to the original problem (16), if the tuning parameters are suitable.
- We recommend selecting them by cross-validation and studying the general shape of the objective function.

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.

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.

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

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

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

References III

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