Weak Identification Robust Methods for Production Function Estimation*

Jorge de la Cal Medina
University of Amsterdam, Tinbergen Institute
j.delacalmedina@uva.nl

November 11, 2025

Abstract

This paper revisits control-function estimation of production functions. We show that, in empirically relevant environments, the structural parameters can be weakly identified even when they are formally point identified. Framing the estimators within a GMM setup, we characterize the implications of weak identification for estimation and inference—most notably, inconsistent estimators and unreliable Wald-type inference. We then discuss and implement identification-robust procedures to conduct inference for control-function estimators. Thereby we provide a guideline for applied work to assess identification and to report inference that remains valid when identification is weak.

*I am grateful for helpful comments from Isaiah Andrews, Giovanni Angelini, Filippo Biondi, Giuseppe Cavaliere, Aureo de Paula, Sabien Dobbelaere, Massimo Giuliodori and Franc Klaaassen and from participants at the MInt Seminar at the University of Amsterdam), the Economics Seminar at the Vrije Universiteit Amsterdam, the Econometrics Seminar at the University of Bologna, the Econometrics Seminar at the University of Zurich, and the 2025 EARIE, EEA, IAAE, and IPDC conferences.

1 Introduction

Production functions are central objects in economics. Their estimation has a long history, yet the econometric difficulties that motivated early work remain central today: firms choose inputs in response to productivity shocks unobserved by the econometrician, rendering inputs endogenous. Control-function methods, put forward by Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), and Ackerberg et al. (2015) (ACF) to address this endogeneity problem, have been used extensively in applied work (see Ackerberg et al. (2007) for a survey).

In this paper, we revisit the identification in the control-function framework, focusing on issues of weak identification. Weak identification arises when parameters are formally point identified but the identifying variation in the data is small, causing standard asymptotic approximations to fail (Lewbel, 2019). We find empirically relevant cases which might give rise to weak identification in the control-function framework, particularly when the productivity proxy has a small little signal relative to noise—a weak proxy.

Our contribution is threefold. First, we cast the control-function estimator into the GMM framework, which allows us to characterize weak identification and show how it leads to inconsistent estimators and nonstandard asymptotic behavior. Second, we discuss various methods to conduct robust inference. Third, we analyze simulated and empirical data where the theoretical predictions are borne out.

The paper proceeds as follows. Section 2 introduces the model and discusses the weak-identification problem. Section 3 develops identification-robust inference procedures. Section 4 reports Monte Carlo experiments based on a dynamic firm model, illustrating the consequences of weak identification and how the proposed robust procedures maintain correct inference. Section 5 applies the methods to data from Chile and the U.S., to show how the quality of the data affects the identification strength, specifically how noisy proxies lead to less reliable estimates. Finally, Section 6 concludes.

2 Identification Analysis

We begin by introducing the control-function approach to the estimation of production functions of OP, LP, and ACF. We specify a parametric version of the model and embed it in a GMM setting. This allows us to formalize identification and characterize weak identification. The parametric specification makes the model more tractable, which is helpful to illustrate the key aspects. However, the same identification issues discussed here arise in the standard semi-parametric setting discussed in OP/LP/ACF.

2.1 Model

We consider a Cobb–Douglas value-added production function for firm i at time t:

$$y_{it} = \beta_k \, k_{it} + \beta_\ell \, \ell_{it} + \omega_{it} + \varepsilon_{it}, \tag{1}$$

where y is (log) output, k and ℓ denote (log) capital and labor, ω is firm-specific productivity, and ε is an idiosyncratic shock. Productivity evolves according to a stationary, linear AR(1) process,

$$\omega_{it} = \rho \, \omega_{it-1} + \xi_{it}, \qquad \rho \in (-1, 1), \tag{2}$$

with productivity innovation ξ. OP/LP/ACF leave the law of motion for productivity unspecified. As mentioned above, we impose a parametric structure to make the model more tractable. Note that ACF also does this to illustrate their estimation procedure. We treat β_k and β_ℓ as the parameters of interest and ρ as a nuisance parameter.

Productivity is unobserved by the econometrician but observed by the firm when choosing inputs. Hence, ω_{it} is correlated with k_{it} and ℓ_{it}, making these inputs endogenous in (1). Following the control-function approach of OP, LP, and ACF, we posit a one-to-one mapping—a control function—between productivity ω_{it} and an observable proxy input, such as intermediate inputs m_{it}.[1] The proxy does not enter (1) directly; otherwise it would not supply independent variation for identification.

Unlike the original OP/LP/ACF implementations, we allow for an error term in the control function to accommodate a noisy proxy. This is in line with econometric treatments of proxy variables (Wooldridge, 2010, Ch. 4). Further, allowing for noise in the control function is theoretically essential for our analysis. It is necessary to study how identification weakens as the informative signal about productivity becomes small relative to the noise component.

To give the error a concrete interpretation, we assume it reflects measurement error in the observed proxy. This assumption is well motivated: in practice, intermediate inputs, such as materials,

[1] We leave the detailed derivation of this mapping to the existing literature (see ACF) and directly posit it here.

are often not directly observed but constructed from expenditure or revenue data, and are therefore measured with considerable noise. This is also noted by ACF who incorporate measurement error in intermediate inputs in their Monte Carlo analysis (as we do in Section 4). Measurement error thus provides a natural and empirically relevant source of proxy error.

We represent the relationship between productivity and the proxy input as linear:

$$\omega_{it} = \delta_k k_{it} + \delta_m m_{it}^*,$$

where m_{it}^* denotes the firm's true proxy input choice. The econometrician observes m_{it}, which is subject to classical measurement error,

$$m_{it} = m_{it}^* + \nu_{it},$$

where $\mathbb{E}[\nu_{it} \mid m_{it}^*] = 0$. Substituting yields the control-function specification

$$\omega_{it} = \delta_k \, k_{it} + \delta_m \, m_{it} + \eta_{it}, \tag{3}$$

where $\eta_{it} := -\delta_m \nu_{it}$ captures the measurement error in the observed proxy. The coefficient δ_m measures the extent to which the proxy conveys information about productivity; as $\delta_m \to 0$, the proxy becomes uninformative and identification weakens.

Equations (1)–(3) describe the structural environment. Identification in the control function approach rests on the timing of firms' decisions. Following OP/LP/ACF we assume that capital is a predetermined (state) variable, while labor and the proxy input are chosen flexibly within the period. Let z_{it} denote the vector of inputs chosen by firm i in period t: $z_{it} = (k_{i,t+1}, \ell_{it}, m_{it})'$. Firms choose inputs z_{it} after observing current productivity ω_{it} but before the realization of the transitory shock ε_{it}. Accordingly, ε_{it} is uncorrelated with both current and past inputs. The innovation to productivity, ξ_{it}, is realized at the beginning of the period and is therefore uncorrelated with past input choices. By construction, η_{it} is mean independent of the firm's information set at the end of period $t-1$, and thus orthogonal to all past input choices. These timing assumptions imply the following sequential exogeneity (orthogonality) conditions:

$$\mathbb{E}[\varepsilon_{it} \mid z_{i\tau}] = 0,\ \tau \leq t \tag{4}$$

$$\mathbb{E}[\xi_{it} \mid z_{i,\tau-1}] = 0,\ \tau \leq t \tag{5}$$

$$\mathbb{E}[\eta_{it} \mid z_{i\tau}] = 0,\ \tau \leq t \tag{6}$$

2.2 Estimation

The previous model can be estimated by GMM. Combining (1) and (2) with the lagged production function yields the one–step equation

$$y_{it} = \beta_k k_{it} + \beta_\ell \ell_{it} + \rho\big(y_{it-1} - \beta_k k_{it-1} - \beta_\ell \ell_{it-1}\big) + \underbrace{\xi_{it} + \varepsilon_{it} - \rho\,\varepsilon_{it-1}}_{u_{it}}. \tag{7}$$

Note that in (7), lagged output y_{it-1} is endogenous because it contains the previous-period transitory shock ε_{it-1}, which also enters the composite error u_{it} through ε_{it-1}. The control-function relation (3) implies that the proxy input m_{it-1} is correlated with past productivity ω_{it-1}, and hence with y_{it-1}, but—under our model assumptions—is uncorrelated with u_{it}. Therefore, m_{it-1} provides a valid instrument for y_{it-1}. Effectively, we have expressed the proxy for productivity as an instrument for output, which naturally allows us to adopt the GMM framework to analyze identification strength.

Let $\theta = (\beta_k, \beta_\ell, \rho)' \in \Theta \subset \mathbb{R}^{d_\theta}$, where Θ is a compact parameter space that contains the true value θ_0 in its interior. Under (4)–(6), any $z_{i\tau}$ with $\tau \leq t-1$ constitutes a valid instrument vector. Since $d_\theta = 3$, we require at least three instruments ($d_z \geq 3$) to achieve identification. A natural choice is $z_{it-1} = \{k_{it-1},\, \ell_{it-1},\, m_{it-1}\}$, which leads to a just-identified model, although additional lags can be included to improve efficiency. Define the moment contribution

$$f_{it}(\theta) := z_{it-1}\Big\{\, y_{it} - \beta_k k_{it} - \beta_\ell \ell_{it} - \rho\big(y_{it-1} - \beta_k k_{it-1} - \beta_\ell \ell_{it-1}\big)\Big\}. \tag{8}$$

We consider a panel of n firms observed for T time periods. Throughout the analysis we let $n \to \infty$ while T is fixed. We assume that $\{(y_{it}, k_{it}, \ell_{it}, m_{it}) : i = 1, \ldots, n,\ t = 1, \ldots, T\}$ are independent across i.

Let $f_i(\theta) := (T-1)^{-1}\sum_{t=2}^{T} f_{it}(\theta)$ and $\bar{f}_n(\theta) := n^{-1}\sum_{i=1}^{n} f_i(\theta)$. Define

$$Q_n(\theta) := n\,\bar{f}_n(\theta)' W_n \bar{f}_n(\theta),$$

for a positive definite weight W_n. The GMM estimator is given by the minimiser

$$\hat{\theta} = \arg\min_{\theta} Q_n(\theta).$$

Under the usual regularity conditions $\hat{\theta}$ is consistent and asymptotically normal (Hansen, 1982).

Efficiency of the estimator depends on the choice of the weighting matrix W_n. The asymptotic covariance matrix of the sample moments is

$$V_{ff}(\theta_0) = \mathrm{Var}\left(\sqrt{n}\,\bar{f}_n(\theta_0)\right),$$

and the optimal weighting matrix is its inverse, $W_n = V_{ff}(\theta_0)^{-1}$. This choice yields the efficient GMM estimator, which is asymptotically efficient within the class of estimators based on the same set of instruments (Hansen, 1982; Newey and McFadden, 1994).

Since $V_{ff}(\theta_0)$ is unknown, it must be estimated. We define the following cluster–robust estimator, evaluated at some θ,

$$\widehat{V}_{ff}(\theta) = \frac{1}{n}\sum_{i=1}^{n}\left(f_i(\theta) - \bar{f}_n(\theta)\right)\left(f_i(\theta) - \bar{f}_n(\theta)\right)'. \tag{9}$$

Under heteroskedasticity and within–firm serial correlation, the estimator in (9) is consistent for $V_{ff}(\theta)$ (Liang and Zeger, 1986; Stock and Watson, 2008).

A practical implementation is the two–step GMM estimator of Hansen (1982). In the first step, a preliminary consistent estimator $\hat{\theta}_1$ is obtained by minimizing the criterion function $Q_n(\theta)$ using an arbitrary positive definite weighting matrix, such as the identity. Using this first-step estimate, the covariance matrix of the sample moments is estimated as $\widehat{V}_{ff}(\hat{\theta}_1)$. The two–step GMM estimator $\hat{\theta}_2$ is then defined as the minimizer of

$$Q_{2,n}(\theta) = n\,\bar{f}_n(\theta)'\widehat{V}_{ff}(\hat{\theta}_1)^{-1}\bar{f}_n(\theta).$$

Under standard regularity conditions, $\hat{\theta}_2$ is asymptotically efficient within the class of estimators that use the same set of instruments (Hansen, 1982; Newey and McFadden, 1994).

An alternative is the continuously updated estimator (CUE) of Hansen et al. (1996), which replaces the fixed weighting matrix with one that depends on the current parameter value. The CUE $\hat{\theta}_c$ is defined as the minimizer of

$$Q_{c,n}(\theta) = n\,\bar{f}_n(\theta)'\widehat{V}_{ff}(\theta)^{-1}\bar{f}_n(\theta).$$

Unlike the two–step estimator, the CUE optimizes over the weighting matrix rather than updating the latter in separate stages. It shares the same asymptotic efficiency as $\hat{\theta}_2$ but often exhibits improved finite–sample performance. Crucially, it forms the basis for identification–robust inference methods discussed in Section 3.

2.3 Identification failure

Identification requires that θ_0 is the unique solution to the population moment condition $\mathbb{E}[f_i(\theta)] = 0$ for $\theta \in \Theta$. A necessary condition for local identification is that the Jacobian of the population moments, evaluated at the true values, $\mathbb{E}\big[\partial f_i(\theta)/\partial\theta'\big]_{\theta=\theta_0}$, has full column rank of $d_\theta = 3$ (Hansen, 1982).

Using (8), the Jacobian at cross-section t, $\mathbb{E}\big[\partial f_{it}(\theta)/\partial\theta'\big]$, is given by

$$\begin{aligned}
&-\Big(\mathbb{E}\big[\partial f_{it}(\theta)/\partial\beta_k\big] \vdots \mathbb{E}\big[\partial f_{it}(\theta)/\partial\beta_\ell\big] \vdots \mathbb{E}\big[\partial f_{it}(\theta)/\partial\rho\big]\Big) \\
&= -\Big(\mathbb{E}\big[z_{it-1}(k_{it} - \rho\, k_{it-1})\big] \vdots \mathbb{E}\big[z_{it-1}(\ell_{it} - \rho\,\ell_{it-1})\big] \vdots \mathbb{E}\big[z_{it-1}(y_{it-1} - \beta_k k_{it-1} - \beta_\ell \ell_{it-1})\big]\Big)
\end{aligned}$$

To illustrate how identification may fail in the control–function framework, we consider a set of stylized but economically meaningful cases that isolate distinct microeconomic sources of non-identification. These cases provide useful benchmarks for understanding how limited input or proxy variation undermines identification. We suppress the firm index i for notational simplicity.

Case (i): Linear labor policy. Consider the case in which labor is chosen according to a linear decision rule,

$$\ell_t = a_{\ell k}\, k_t + a_{\ell\omega}\,\omega_t,$$

for some constants $a_{\ell k}$ and $a_{\ell\omega}$. This is related to the "functional dependence problem" discussed by ACF, who note that under the assumptions of OP/LP/ACF, labor is a deterministic function of capital and productivity. Under linear policy and using (2), we obtain $-\mathbb{E}[\partial f_{it}(\theta)/\partial\beta_\ell] = a_{\ell k}\,\mathbb{E}\big[z_{t-1}(k_t - \rho\, k_{t-1})\big]$. Hence, the first and the second column of the Jacobian are proportional to each other.

Case (i): Static capital and uninformative proxy. A second case of underidentification arises when capital does not adjust within the observed periods. This situation was emphasized by LP, who note that capital typically evolves only sluggishly over time. If, in addition, the proxy variable carries no independent signal about productivity, the model parameters cannot be identified. Suppose capital is fixed over time, $k_t = k_{t-1}$, and the proxy is uninformative, that is, $\delta_m = 0$ in (3). In this case, $-\mathbb{E}[\partial f_{it}(\theta)/\partial\beta_k] = (1-\rho)\mathbb{E}[z_{it-1}k_{t-1}]$ and $-\mathbb{E}[\partial f_{it}(\theta)/\partial\rho] = \delta_k\mathbb{E}[z_{it-1}k_{t-1}]$. Hence, the first and the third column of the Jacobian are proportional to each other.

Either of these two cases leads to a rank-deficient Jacobian. Both cases taken together constitute an extreme case of underidentification since then all three columns are proportional to each other. Although they are unlikely to hold exactly in practice, their empirical counterparts are common: first, labor adjustment is often approximately proportional to capital, second, capital stocks evolve sluggishly and proxy variables may convey only weak information about productivity. Taken together, these stylized cases provide useful intuition for how weak identification can arise in the control–function framework. In Section 4, we analyze these cases using simulated data calibrated to a firm-level model. But before that, we analyze the asymptotic behavior of the estimator under weak identification.

2.4 Weak Identification

We show that the estimator $\hat{\beta}$ cannot be consistently estimated under weak identification. In order to formalize weak identification, we adopt the partial weak identification framework of Stock and Wright (2000). We consider a sequence of data–generating processes indexed by n under which the rank of the Jacobian of the population moment conditions becomes arbitrarily close to being reduced. Specifically, a subset of the Jacobian's columns is treated as local to zero, and the corresponding parameters are only weakly identified, while the remaining parameters retain standard identification.

Let $\theta = (\beta', \rho)'$, where $\beta = (\beta_k, \beta_\ell)'$. Define the population moments as

$$m_n(\beta,\rho) := \mathbb{E}\big[\bar{f}_n(\beta,\rho)\big].$$

Using an exact Taylor expansion of second degree, we can express them as

$$m_n(\beta, \rho) = m_n(\beta_0, \rho_0) + M_{\beta,n}(\rho)(\beta - \beta_0) + (\rho - \rho_0)\, m_{\rho,n}(\beta_0),$$

where $M_{\beta,n}(\rho) = \mathbb{E}\left[\partial \bar{f}_n(\beta, \rho)/\partial \beta'\right]$ and $m_{\rho,n}(\beta) = \mathbb{E}\left[\partial \bar{f}_n(\beta, \rho)/\partial \rho\right]$.

Assumption 2.1 (Partial local-to-zero Jacobian)**.** Along a sequence of DGPs indexed by n,

(i) $M_{\beta,n}(\rho) = \frac{1}{\sqrt{n}}\, C(\rho)$ where $C(\rho) \in \mathbb{R}^{d_z \times 2}$ is continuous in ρ and bounded on Θ.

(ii) $m_{n,\rho}(\beta_0) = r$ where $r \in \mathbb{R}^{d_z \times 1}$.

(iii) The matrix $[C(\rho)\ r] \in \mathbb{R}^{d_z \times 3}$ has full rank, uniformly in ρ.

We now show the key result that $\hat{\beta}$ is inconsistent. For this we follow Stock and Wright (2000). We impose a set of regularity conditions that are standard and hold under mild assumptions. Define the centered empirical process on the parameter space Θ as

$$\psi_n^f(\theta) \ :=\ \sqrt{n}\left\{\bar{f}_n(\theta) - m_n(\theta)\right\}.$$

Assumption 2.2 (Weak convergence)**.** The process ψ_n^f converges weakly to a mean–zero Gaussian process ψ_0^f:[2]

$$\psi_n^f \Rightarrow \psi_0^f.$$

Assumption 2.3 (Positive definite weighting matrix)**.** The weighting matrix W_n is symmetric and positive definite for all n, and satisfies

$$W_n \xrightarrow{p} W_0,$$

where W_0 is symmetric and positive definite.

Before analyzing the weakly identified component, we first establish that $\hat{\rho}$ converges at the usual $\sqrt{n}$ rate.

Proposition 2.4 ($\sqrt{n}$-consistency of the strongly identified parameter)**.** *Let Assumptions 2.1–2.3 hold. Then, for any GMM minimizer $\hat{\theta} = (\hat{\beta}', \hat{\rho})'$ of $Q_n(\theta) = n\, \bar{f}_n(\theta)' W_n \bar{f}_n(\theta)$, $\hat{\rho}$ is $\sqrt{n}$-consistent:*

$$\hat{\rho} \ \xrightarrow{p}\ \rho_0 \ \textit{and}\ \sqrt{n}\,(\hat{\rho} - \rho_0) \ =\ O_p(1).$$

[2] $\Rightarrow$ denotes weak convergence of random functions on Θ with respect to the supremum norm (see van der Vaart, 2000, Sec. 18)

Proof. By Assumption 2.1, and since $m_n(\beta_0, \rho_0) = 0$

$$m_n(\beta, \rho) = \frac{C(\rho)}{\sqrt{n}}(\beta - \beta_0) + (\rho - \rho_0)\, r$$

Hence,

$$\frac{1}{n} Q_n(\beta, \rho) \xrightarrow{p} (\rho - \rho_0)^2 r' W_0 r.$$

By the positive definiteness of W_0 (by Assumption 2.3), and since $(\rho - \rho_0)r = 0$ only if $\rho = \rho_0$ (by Assumption 2.1), consistency of $\hat{\rho}$ for ρ_0 follows.

For $\sqrt{n}$–consistency, decompose

$$Q_n(\theta) = \left(\psi_n^f(\theta) + \sqrt{n}\, m_n(\theta)\right)' W_n \left(\psi_n^f(\theta) + \sqrt{n}\, m_n(\theta)\right).$$

Since $\hat{\theta}$ minimizes $Q_n(\theta)$ we can write

$$\begin{aligned} Q_n(\hat{\theta}) - Q_n(\theta_0) = & \left(\psi_n^f(\hat{\theta})\right)' W_n \left(\psi_n^f(\hat{\theta})\right) \\ & + 2\left(\psi_n^f(\hat{\theta})\right)' W_n \left(\sqrt{n}\, m_n(\hat{\theta})\right) \\ & + \left(\sqrt{n}\, m_n(\hat{\theta})\right)' W_n \left(\sqrt{n}\, m_n(\hat{\theta})\right) \\ & - \left(\psi_n^f(\theta_0)\right)' W_n \left(\psi_n^f(\theta_0)\right) \leq 0 \end{aligned}$$

Equivalently,

$$(\sqrt{n}\,(\hat{\rho} - \rho_0))^2\, r' W_n r + 2\sqrt{n}\,(\hat{\rho} - \rho_0)\, r' W_n \big(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)\big) + d_{1n} \;\leq\; 0.$$

where

$$d_{1n} := \left(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)\right)' W_n \left(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)\right) - \left(\psi_n^f(\theta_0)\right)' W_n \left(\psi_n^f(\theta_0)\right)$$

We can bound

$$\begin{aligned} (\sqrt{n}\,(\hat{\rho} - \rho_0))^2\, r' W_n r \;&\geq\; \lambda_{\min}(W_n) \left\|\sqrt{n}\,(\hat{\rho} - \rho_0)\, r\right\|^2, \\ \sqrt{n}\,(\hat{\rho} - \rho_0)\, r' W_n \big(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)\big) \;&\geq\; -\left\|\sqrt{n}\,(\hat{\rho} - \rho_0)\, r\right\| \left\|W_n \big(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)\big)\right\|. \end{aligned}$$

which implies

$$\lambda_{\min}(W_n)\left\|\sqrt{n}\,(\hat{\rho}-\rho_0)\,r\right\|^2-\left\|\sqrt{n}\,(\hat{\rho}-\rho_0)\,r\right\|\,\left\|W_n\big(\psi_n^f(\hat{\theta})+C(\hat{\rho})(\hat{\beta}-\beta_0)\big)\right\|+d_{1n}\leq 0.$$

Since $\lambda_{\min}(W_n)>0$,

$$\left\|\sqrt{n}\,(\hat{\rho}-\rho_0)\,r\right\|\;\leq\;\frac{\left\|W_n\big(\psi_n^f(\hat{\theta})+C(\hat{\rho})(\hat{\beta}-\beta_0)\big)\right\|+\sqrt{\left\|W_n\big(\psi_n^f(\hat{\theta})+C(\hat{\rho})(\hat{\beta}-\beta_0)\big)\right\|^2-4\,\lambda_{\min}(W_n)\,d_{1n}}}{2\,\lambda_{\min}(W_n)}.$$

Moreover, under Assumptions 2.2-2.3, the right-hand side of the inequality above is $O_p(1)$, implying

$$\left\|\sqrt{n}\,(\hat{\rho}-\rho_0)\,r\right\|=O_p(1).$$

Since $r\neq 0$ (by Assumption 2.1), $\sqrt{n}$–consistency of $\hat{\rho}$ follows.

□

By Proposition 2.4, the estimator $\hat{\rho}$ converges at the standard $\sqrt{n}$ rate. It thus suffices to consider local $\sqrt{n}$-perturbations when characterizing the large-sample behavior of the GMM objective function. Define $\tilde{\rho}=\rho_0+\lambda/\sqrt{n}$ for some scalar λ, so that $\tilde{\rho}\xrightarrow{p}\rho_0$ and $\sqrt{n}\,(\tilde{\rho}-\rho_0)=O_p(1)$, as stated in Proposition 2.4.

Proposition 2.5 (Consequences of weak identification)**.** *Suppose that Assumptions 2.1-2.3 hold. Then the limiting process for the GMM objective function allows for the following representation*

$$Q_n(\beta,\tilde{\rho})\Rightarrow\big(\psi_0^f(\beta,\rho_0)+C(\rho_0)(\beta-\beta_0)+\lambda r\big)'\,W_0\,\big(\psi_0^f(\beta,\rho_0)+C(\rho_0)(\beta-\beta_0)+\lambda r\big).$$

Proof. Under Assumption 2.1,

$$\sqrt{n}\,\bar{f}_n(\beta,\tilde{\rho})=\psi_n^f(\beta,\tilde{\rho})+C(\tilde{\rho})(\beta-\beta_0)+\lambda\,r.$$

By Assumption 2.2, $\psi_n^f(\beta,\tilde{\rho})\Rightarrow\psi_0^f(\beta,\rho_0)$. By Assumption 2.3, $W_n\xrightarrow{p}W_0$. Applying Slutsky we obtain the stated result. □

Proposition 2.5 parallels the result in Stock and Wright (2000, Theorem 1). Because $C(\rho)$ is bounded, $Q_n(\beta,\tilde{\rho})$ is stochastically bounded uniformly in β. Hence, we cannot consistently estimate

β. and $\hat{\beta}$ is only $O_p(1)$ with a non-normal limiting distribution. The behavior of the strongly identified parameter ρ is also affected. Although $\hat{\rho}$ is consistent, its limiting distribution is nonnormal because it depends on the random limit of $\hat{\beta}$.

Intuitively, weak identification manifests in the shape of the sample criterion: its curvature in the β direction is nearly flat, leading to ridges or plateaus along which many values of β almost minimizes the criterion. The estimator $\hat{\beta}$ therefore wanders stochastically over this set, while $\hat{\rho}$ adjusts conditionally on it. These features explain the heavy-tailed, asymmetric sampling behavior of GMM estimators documented in simulation studies and formalized by Stock and Wright (2000). We will show this in Section 4 using simulated data from a firm optimization model.

2.5 Proxy strength and a concentration parameter

The weak–proxy problem in our setting closely parallels the weak–instrument problem in IV models. In linear IV models, identification strength is summarized by the *concentration parameter* (Stock et al., 2002), which measures the strength of the relationship between the endogenous regressor and the instrument. When the concentration parameter is small, the IV estimator has nonstandard asymptotics, first–stage F-statistics are low, and inference based on Wald tests becomes unreliable.

Although our model is nonlinear, we can construct an analogue of the IV concentration parameter to summarize the strength of the proxy. This analogue should be viewed as a conservative measure of identification strength: the nonlinear structure can only strengthen identification relative to this linearized benchmark.

Recall from (7) that y_{t-1} enters the structural equation as an endogenous regressor: it is correlated with u_t through the term ε_{t-1}. At the same time, under the control-function relationship (3), the proxy input m_{t-1} is informative about past productivity ω_{t-1} and hence about y_{t-1}, while remaining orthogonal to u_t under (4)–(6). Therefore, m_{t-1} plays the role of an instrument for y_{t-1}.

To make the analogy with IV precise, consider the (linearized) reduced form for y_{t-1}. Combining (1) and (3) one period lagged gives

$$y_{t-1} = \beta_k k_{t-1} + \beta_\ell \ell_{t-1} + \delta_k k_{t-1} + \delta_m m_{t-1} + (\varepsilon_{t-1} + \eta_{t-1}),$$

so that, conditional on (k_{t-1}, ℓ_{t-1}), the proxy m_{t-1} shifts y_{t-1} with slope δ_m. In this sense, δ_m plays the same role as the first-stage coefficient on the excluded instrument in a standard IV regression, and $\varepsilon_{t-1} + \eta_{t-1}$ plays the role of the first-stage error.

The analogue of the IV concentration parameter at cross-section t is then

$$\mu_t^2 \;=\; \frac{n\,\delta_m^2\,\mathrm{Var}(m_{t-1} \mid k_{t-1}, \ell_{t-1})}{\sigma_\varepsilon^2 + \sigma_\eta^2}. \tag{10}$$

A larger μ_t^2 implies stronger identification. It is straightforward to see that μ_t^2 is strictly decreasing in the proxy noise variance σ_η^2, and increasing in $|\delta_m|$, given $\mathrm{Var}(m_{t-1} \mid k_{t-1}, \ell_{t-1})$. Thus, a less informative proxy or tighter collinearity between the proxy and inputs weakens identification.

As a case in point, suppose m_t is subject to classical measurement error. Keeping $\mathrm{Var}(m_{t-1} \mid k_{t-1}, \ell_{t-1})$ fixed, an increase in measurement error simultaneously reduces $|\delta_m|$ and raises σ_η^2, thereby lowering μ_t^2. This mirrors the well-known result in linear IV models where measurement error leads to weak instruments.

In summary, μ_t^2 serves as a diagnostic for proxy strength in the control-function approach. When μ_t^2 is small, the proxy is effectively weak: the GMM estimator inherits the weak-identification behavior described above (slow convergence of $\hat{\beta}$, non-Gaussian limits, and unreliable Wald inference). Section 4 shows this mechanism in Monte Carlo simulations calibrated to a firm optimization model.

3 Identification Robust Methods

We employ an identification-robust test statistic to conduct inference on the production function parameters within the control function approach. Conventional inference methods, such as Wald statistics and t-statistics, rely on asymptotic normality, which breaks down under weak identification, resulting in nonstandard distributions and size distortions (Staiger and Stock, 1997; Stock and Wright, 2000). Selecting instruments or proxy variables based on standard pre-testing, such as the commonly used F-test, does not resolve the issue, as it might introduce selection bias resulting in a test that is not size-controlled.

Identification-robust statistics overcome these problems by maintaining a valid limiting distribution, ensuring reliable hypothesis testing and confidence intervals (Andrews and Stock, 2005). These statistics are defined for the CUE.

3.1 Bootstrap pre-test for proxy relevance

Angelini et al. (2024) propose a bootstrap normality diagnostic that makes it possible to assess a proxy's relevance. The null hypothesis is that all proxies (and instruments) are strong. Let $\hat{\theta}_n$ denote the CUE obtained from the full sample and let V_θ be a consistent estimate of its asymptotic covariance, which we can compute from $V_{ff}(\hat{\theta})$. Let $\hat{\theta}_n^*$ be the bootstrapped counterpart, using Moving Block Bootstrap (MBB) technique. The MBB is similar in spirit to a standard bootstrap with replacement. However, instead of resampling one observation at a time the MBB resamples blocks of observations, in out case at the firm-level, in order to replicate their serial dependence structure (Angelini et al., 2024). We can then form

$$\Gamma_n^* = \sqrt{n}\, V_\theta^{-1/2}\big(\hat{\theta}_n^* - \hat{\theta}_n\big).$$

Under the null, Γ_n^* is asymptotically standard normal. Testing multivariate normality—e.g. with the Doornik–Hansen or a set of Kolmogorov–Smirnov statistics—therefore provides a valid decision rule. Failure to reject normality implies that standard Wald or CUE inference remains appropriate; rejection signals weak proxies, in which case one should switch to identification-robust procedures such as the AR-GMM confidence sets introduced above.

The pre-test does not distort subsequent inference because the bootstrap normality statistic is asymptotically independent of any post-test identification-robust statistic. It is also straightforward to implement: aside from the re-estimations needed for the bootstrap, no extra first-stage regressions or critical-value adjustments are required, and the method accommodates multiple proxies, conditional heteroskedasticity, and zero-censored instruments. For further technical details and simulation evidence, see Angelini et al. (2024).

3.2 AR-GMM test

A straightforward identification-robust statistic is the GMM extension of the AR test (Anderson and Rubin, 1949), the AR-GMM. It is based on the S-statistic of Stock and Wright (2000). Let $\theta \in \mathbb{R}^{d_\theta}$ be the parameter of interest and $f_n(\theta) \in \mathbb{R}^{d_f}$ the sample moments with which it is estimated. To test the null hypothesis $H_0 : \theta_0 = \theta^*$, the S-statistic is equivalent to the objective function of the CUE evaluated at θ^*.

$$S(\theta^*) = n f_n(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} f_n(\theta^*) \tag{11}$$

The S-statistic converges under H_0 to a $\chi^2_{d_f}$, irrespective of the identification strength (see Theorem 2 in Stock and Wright (2000)).

The AR-GMM test has power against the alternative hypothesis $\theta \neq \theta^*$ and against the violation of moment restrictions in overidentified models. Its power diminishes as the number of moment restrictions increases, since the degrees of freedom of its limiting distribution correspond to the number of moment conditions. Consequently, when the number of instruments substantially exceeds the number of structural parameters, the statistic exhibits low power. To address this limitation, alternative, more powerful identification-robust test statistics have been developed (Moreira, 2003; Kleibergen, 2002, 2005). Future research could build on these advancements to formulate similar tests adapted to the control function approach.

3.3 Robust Confidence Sets

From Proposition 2 it follows that the $100 \times (1-\alpha)\%$ confidence set for θ is

$$CS(\alpha) = \{\theta^* : S(\theta^*) \leq \chi^2_{d_f}(1-\alpha)\} \tag{12}$$

where $\chi^2_{d_f}(\alpha)$ represents the $100 \times (1-\alpha)$th percentile of the limiting distribution of $\chi^2_{d_f}$. Since these tests are not quadratic functions of θ^*, they cannot be directly inverted to obtain the confidence set. Thus, the confidence sets are not expressed as an estimator $\pm$ a multiple of standard error. Instead, one must compute $S(\theta^*)$ for each value on a grid of θ^* values to check if it falls within $CS(\alpha)$.

Generally, the confidence set based on the S-statistic can take different forms. Bounded sets typically indicate strong parameter identification and are ideally convex. Bounded but concave sets, however, may still point to issues related to weak identification. Unbounded confidence sets often result from weak identification, suggesting that the model is nearly underidentified. Empty sets indicate a failure to satisfy the moment conditions, pointing to possible model misspecification. This last type of confidence set will be discussed further in the next subsection, where we introduce a test for overidentifying restrictions. Wang and Zivot (1998) discuss different shapes of bivariate confidence sets for the AR statistic in the presence of weak instruments. Specifically, these sets may take on elliptical or hyperbolic shapes, with the latter indicating potential issues with weak identification.

Confidence sets for a subvector of θ can be derived by substituting the remaining elements of

θ with their estimates. The S-statistic is then computed over a grid of θ values, varying only the elements of the subvector in question while keeping the other vector elements fixed. An alternative approach is to first construct a valid confidence set for θ and then extract the relevant subset via projection (Dufour, 1997, Sec. 5.2), which typically yields an asymptotically conservative confidence set.

3.4 Subset AR-GMM test

Let $\theta = (\theta_1, \theta_2)$ with parameter of interest $\theta_1 \in \mathbb{R}^{d_{\theta_1}}$ and $\theta_2 \in \mathbb{R}^{d_{\theta_2}}$ is treated as nuisance. To test

$$H_0 : \ \theta_1 = \theta_1^*,$$

form the *subset* S–statistic by concentrating out θ_2 via the restricted CUE:

$$\hat{\theta}_2(\theta_1^*) \ := \ \arg\min_{\theta_2} \ S(\theta_1, \theta_2)\big|_{\theta_1 = \theta_1^*}.$$

The test statistic is then $S(\theta_1^*, \hat{\theta}_2(\theta_1^*))$.

Under H_0 the limiting distribution of $S(\theta_1^*, \hat{\theta}_2(\theta_1^*))$ is stochastically dominated by $\chi^2_{d_f - d_{\theta_2}}$, with equality when the unrestricted nuisance θ_2 is well identified. Hence using the $\chi^2_{d_f - d_{\theta_2}}$ critical value yields a uniformly valid (large-sample) test that remains size-correct irrespective of the identification strength of θ_2 (Kleibergen and Mavroeidis, 2009, Thm. 2).

An identification-robust $(1 - \alpha) \times 100\%$ confidence set for θ_1 is obtained by inversion:

$$CS(\alpha) \ := \ \Big\{ \theta_1 \in \mathbb{R}^{d_{\theta_1}} : \ S\big(\theta_1, \hat{\theta}_2(\theta_1)\big) \ \leq \ \chi^2_{d_f - d_{\theta_2}}(1 - \alpha) \Big\}.$$

Computation proceeds by gridding θ_1 and re-solving the restricted CUE at each grid point (as in the AR–GMM sets above). As with full-parameter sets, shapes may be bounded, unbounded, or disconnected under weak identification.

Subset tests are (weakly) more powerful than projection-based procedures: non-rejection by the subset AR-GMM test implies non-rejection by its projection counterpart, so projection-based confidence sets are typically more conservative (Dufour, 1997).

4 Monte Carlo Experiment

Our Monte Carlo design follows that of ACF closely. It builds on the analytically solvable dynamic investment model Van Biesebroeck (2003). The parameters are calibrated to reproduce key moments in the Chilean plant data analyzed by Levinsohn and Petrin (2003).

4.1 Setting

Output is produced with a Leontief production function in materials:

$$Y_{it} = \min\Big\{\alpha\, K_{it}^{\beta_k} L_{it}^{\beta_l} \exp\big(\omega_{it}\big),\ \beta_m\, M_{it}^{*}\Big\}\ \exp\big(\varepsilon_{it}\big) \tag{13}$$

with $(\alpha, \beta_k, \beta_l, \beta_m) = (1,\ 0.4,\ 0.6,\ 1)$. Y_{it} denotes the output of firm i in period t; K_{it}, L_{it} and M_{it}^{*} are, respectively, the levels of capital, labour and material inputs. Note that the material inputs are latent since their observations will be ridden with measurement error. ε_{it} is normal with mean zero with standard deviation 0.01. ω_{it} is assumed to follow an AR(1) with auto-regressive coefficient $\rho = 0.7$.

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it} \tag{14}$$

ε_{it} is independent to the current information set and mean-zero. The variance of the normally distributed innovation ξ_{it} (σ_{ξ}^{2}) and the initial value ω_{i0} ($\sigma_{\omega_{i0}}^{2}$) are set such that the standard deviation of ω_{it} is equal to 0.3.

The capital stock evolves as

$$K_{it} \ = \ (1-\delta)\, K_{it-1} + I_{it-1}.$$

where I_{it} denotes capital investment, and depreciation rate is set to $\delta = 0.8$. Firms choose investment to maximise the expected, discounted value of future profits subject to convex capital-adjustment costs. Labor and material inputs are selected contemporaneously with output in period t.

Because materials are used in a fixed proportion to output, the "structural value-added" regression equation (see Gandhi et al. (2020)) is

$$y_{it} = \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it} \tag{15}$$

where lowercase letters denote the logs. The Leontief specification implies that in optimality we have that both arguments of the min-function have to be equal. This gives us the intermediate demand function.

$$m_{it}^* = -\log(\beta_m) + \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it}, \tag{16}$$

where lower case letters denote the logs. We allow for independent measurement error in m_{it}^*.

$$m_{it} = m_{it}^* + \nu_{it}, \tag{17}$$

where ν_{it} is normal with zero mean and standard deviation σ_ν. We will generate data for different values of σ_ν, which adds noise to the control function. For high values of σ_ν we have weak identification, as will be shown in the results.

As ACF point out, the firm's choice on labor input L_{it} is functionally dependent on, or, in the case of measurement error, strongly correlated with, K_{it} and M_{it}. While this dependence can be accommodated in the second-stage moments, it leaves empirical identification more fragile. To strengthen identification, ACF add independent optimization error to the labor input.

$$l_{it} = l_{it}^* + \xi_{it}^l$$

where ξ_{it}^l follows a normal distribution with mean zero and a standard deviation of σ_{ξ_l}.

4.2 Simulation Results

We consider two DGPs that differ only in the variance of the labor optimization error. The first replicates ACF's DGP with $\sigma_{\xi_l} = 0.37$ (see the description of DGP2 in ACF, Section 5); the second is identical except that the labor optimization error is reduced to $\sigma_{\xi_l} = 0.10$. Effectively this reduces the optimization error in l from around 10% to 5%. This reduction will make the consequences of weak identification more salient at lower measurement error, which serves for illustration. For both DGPs we add classical measurement error in the measured materials input of $0\%, 10\%, 20\%, 50\%$, where a 10% measurement error raises the variance of observed materials by 10%.

For each simulation, we estimate (β_l, β_k) using the estimation techniques of LP and ACF, and our one-step CUE estimator from Section 5. We report the conventional (non-robust) Wald test and the identification-robust AR-GMM test from Section 3 for the null at the true parameter vector. In addition, we apply the bootstrap normality test of Angelini et al. (2024), also described in from Section 3, which assesses joint normality of the estimators using Shapiro–Wilk–based diagnostics.

Table B.1 reports Monte Carlo estimates under the first DGP ($\sigma_{\xi_l} = 0.37$). With no measurement error, all three estimators (ACF, LP, CUE) are tightly centered at the true values $(0.6, 0.4)$ and exhibit small dispersion. As the error variance increases, LP shows a clear location drift—estimates for β_l move upward while those for β_k move downward—whereas ACF remains comparatively stable, with only a mild upward drift in β_l and negligible movement in β_k. CUE stays closest to unbiased across all error levels, with only moderate increases in variance. Figure C.1 displays the empirical densities of the simulated estimates with normal overlays. It shows that the empirical densities of all estimates are close to normal at 0% error and remain near–normal as the error increases. This is also reflected when testing. Table B.2 shows that rejection frequencies at the 5% level are close to nominal for both the conventional Wald test and the identification-robust AR–GMM test across all error levels; the Doornik–Hansen joint-normality test rejects somewhat more often as measurement error grows, indicating mild departures from normality in this design. It points toward the normality test being conservative.

Table B.3 reports Monte Carlo estimates under the second DGP ($\sigma_{\xi_l} = 0.10$). With no measurement error, all three estimators (ACF, LP, CUE) are centered at the true values and exhibit small dispersion. As the error variance increases to 10%–20%, identification weakens: LP becomes unstable—estimates for β_l move toward one while those for β_k reach implausible magnitudes pointing toward instability—whereas ACF deteriorates more gradually, with β_l drifting up and β_k drifting down and occasional extreme realizations; CUE remains the most stable over this range with a gradual increase in variance. At 50% error, LP exhibits boundary pile–ups and ACF produces explosive draws in a nontrivial share of replications, while CUE continues to yield interpretable coefficients, albeit with thicker tails and higher dispersion. Figure C.2 displays the empirical densities of the simulated estimates. At 0% error the densities are approximately normal and centered at the truth; as measurement error rises, LP's densities shift (right for β_l, left for β_k) and become spiky near the boundary, ACF develops heavier tails with the same directional drift. While CUE is the most stable, at 20% measurement error and above its sampling distribution shows clear deviations from normality—visible skewness and heavy tails—despite estimates remaining centered. Consistent with these distributional changes, Table B.4 shows that the conventional Wald test increasingly over-rejects invalidating inference; whereas the identification-robust AR–GMM test maintains rejection frequencies close to 5% across error levels. The Doornik–Hansen joint-normality rejection rate rises sharply with measurement error, consistent with the pronounced departures from normality in this design and indicating that the diagnostic is appropriately sensitive.

5 Empirical Analysis

We replicate the empirical exercise in Raval (2023).[3] Specifically, we estimate a Cobb–Douglas revenue production function with capital, labor, and materials,

$$y_{it} \;=\; \beta_k k_{it} \;+\; \beta_l \ell_{it} \;+\; \beta_m m_{it} \;+\; \omega_{it} \;+\; \varepsilon_{it},$$

where productivity follows an AR(1) process,

$$\omega_{it} \;=\; \rho\, \omega_{i,t-1} \;+\; \xi_{it}.$$

The proxy/control function $\Psi(k, \ell, m)$ is approximated by a third–order polynomial in (k, ℓ, m). We implement the ACF control–function estimator in the same specifications as Raval (2023) and re–estimate the same model with our one–step CUE estimator.

5.1 Data

We estimate the model on two sources from Chile and from the U.S. that differ in measurement quality. The Chilean data (ENIA) are plant–level and survey-based. Materials are observed directly as intermediate consumption (raw materials plus energy) and are cleanly separated from wages. The U.S. data (Compustat) are firm–level accounting data; materials must be proxied by COGS − XLR, with COGS mixing materials, direct labor, and overhead and XLR (labor expenses) coverage varying across firms and time. Accordingly, the proxy–state mapping is tighter in Chile and noisier in the U.S., increasing the scope for weak identification in the latter; this pattern is borne out in the estimates below.

To obtain comparable cross–country samples given Raval (2023)'s aggregation (3–digit ISIC for Chile; 2–digit NAICS for the U.S.), we select metals industries at matching breadth: ISIC 381

[3] Data construction details and industry definitions are in App. A of Raval (2023).

in Chile (fabricated metal products, excluding machinery and equipment) and NAICS 33 in the U.S. (manufacturing, including primary and fabricated metals). The resulting panels contain about 4,000 Chilean plant–year observations (1979–1996) and about 8,000 U.S. firm–year observations (1970–2010).

5.2 Chile (ENIA)

Table B.5 shows the results for the Chilean data. It shows close agreement between ACF and CUE for the labor, capital, and materials elasticities, with returns to scale near constant. The non-robust Wald intervals are similar in width and location to the identification-robust Subset-S intervals and remain relatively tight, indicating that identification is reasonably strong in this sample. The bootstrap normality test nevertheless rejects for all coefficients, so inference based on normal approximations should be treated with caution. However, in this case choosing the robust intervals comes at little costs, since they are strongly aligned with the nonrobust ones. Overall, the results are consistent with the cleaner materials measurement in ENIA—plant-level intermediate consumption which plausibly leads to a strong signal from the proxy.

5.3 U.S. (Compustat)

Table B.6 shows the results for U.S. data. Here, the CUE aligns less closely with ACF, especially for capital and in the implied returns to scale, and the identification-robust intervals widen substantially; for the labor elasticity the robust set is unbounded on one side, which is evidence for weak identification. Normality is again rejected across coefficients. These features accord with the noisier proxy in Compustat data, which weakens the mapping from the proxy to productivity and manifests as large, asymmetric, and in places unbounded robust confidence sets.

6 Conclusion

This paper studies weak identification in control-function estimation of production functions. We argue that popular techniques Olley and Pakes (1996); Levinsohn and Petrin (2003); Ackerberg et al. (2006)—may suffer from weak identification when proxy variables have limited explanatory power, yielding biased estimates and unreliable inference and paralleling the weak-instruments problem in IV.

We revisit production-function estimation with control functions when the proxy for productivity carries limited signal. We replace the latent shock with a researcher-feasible index based on observables and impose a forecast-sufficiency restriction that preserves the usual moment conditions while making explicit when a single-index control is adequate. Within this framework we provide a semiparametric identification characterization via an orthogonalized Jacobian, which highlights how weak proxies translate into near-singular information and non-Gaussian sampling behavior.

Building on these primitives, we deliver inference methods that remain valid irrespective of identification strength. A simple bootstrap diagnostic helps flag departures from normality associated with weak signal (Angelini et al., 2024), and identification-robust AR-GMM tests (Stock and Wright, 2000)—implemented using a continuously updated GMM objective with cluster-robust variance—offer reliable inference when conventional Wald procedures do not.

We recommend reporting the normality diagnostic together with identification-robust confidence sets alongside conventional intervals in control-function applications, especially where proxies are constructed from accounting aggregates.

References

Ackerberg, D. A., Caves, K., and Frazer, G. (2006). Structural identification of production functions. MPRA Paper from University Library of Munich, Germany.

Ackerberg, D. A., Caves, K., and Frazer, G. (2015). Identification properties of recent production function estimators. *Econometrica*, 83(6):2411–2451.

Ackerberg, D. A., Lanier Benkard, C., Berry, S., and Pakes, A. (2007). Chapter 63 econometric tools for analyzing market outcomes. volume 6 of *Handbook of Econometrics*, pages 4171–4276. Elsevier.

Anderson, T. W. and Rubin, H. (1949). Estimation of the parameters of a single equation in a complete set of stochastic equations. *The Annals of Mathematical Statistics*, 21:570–582.

Andrews, D. W. and Stock, J. H. (2005). Inference with weak instruments. Technical Working Paper 313, National Bureau of Economic Research, Cambridge, MA.

Angelini, G., Cavaliere, G., and Fanelli, L. (2024). An identification and testing strategy for proxy-svars with weak proxies. *Journal of Econometrics*, 238(2):105604.

Dufour, J.-M. (1997). Some impossibility theorems in econometrics with applications to structural and dynamic models. *Econometrica*, 65(6):1365–1387.

Gandhi, A., Navarro, S., and Rivers, D. A. (2020). On the identification of gross output production functions. *Journal of Political Economy*, 128(8). Electronically published.

Hansen, L. P. (1982). Large sample properties of generalized method of moments estimators. *Econometrica*, 50(4):1029–1054.

Hansen, L. P., Heaton, J., and Yaron, A. (1996). Finite-sample properties of some alternative gmm estimators. *Journal of Business & Economic Statistics*, 14(3):262–280.

Kleibergen, F. (2002). Pivotal statistics for testing structural parameters in instrumental variables regression. *Econometrica*, 70(5):1781–1803.

Kleibergen, F. (2005). Testing parameters in gmm without assuming that they are identified. *Econometrica*, 73:1103–1123.

Kleibergen, F. and Mavroeidis, S. (2009). Inference on subsets of parameters in GMM without assuming identification. Working paper (updated version), Brown University.

Levinsohn, J. A. and Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–340.

Lewbel, A. (2019). The identification zoo: Meanings of identification in econometrics. *Journal of Economic Literature*, 57(4):835–903.

Liang, K.-Y. and Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, 73(1):13–22.

Moreira, M. J. (2003). A conditional likelihood ratio test for structural models. *Econometrica*, 71(4):1027–1048.

Newey, W. K. and McFadden, D. (1994). Large sample estimation and hypothesis testing. In Engle, R. F. and McFadden, D., editors, *Handbook of Econometrics*, volume 4, chapter 36, pages 2111–2245. Elsevier.

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

Raval, D. (2023). Testing the Production Approach to Markup Estimation. *The Review of Economic Studies*, 90(5):2592–2611.

Staiger, D. and Stock, J. H. (1997). Instrumental variables regression with weak instruments. *Econometrica*, 65(3):557–586.

Stock, J. H. and Watson, M. W. (2008). Heteroskedasticity-robust standard errors for fixed effects panel data regression. *Econometrica*, 76(1):155–174.

Stock, J. H. and Wright, J. H. (2000). Gmm with weak identification. *Econometrica*, 68(5):1055–1096.

Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518–529.

Van Biesebroeck, J. (2003). Productivity dynamics with technology choice: An application to automobile assembly. *Review of Economic Studies*, 70(1):167–198.

Wang, J. and Zivot, E. (1998). Inference on structural parameters in instrumental variables regression with weak instruments. *Econometrica*, 66(6):1389–1404.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA, 2nd edition.

Appendices

A Appendix A Monte Carlo Design

We simulate $n = 1000$ firms observed for $T = 10$ consecutive periods. Unless otherwise noted, parameter values are taken from ACF. The parameters are calibrated so that roughly 95% of the cross-sectional variation in capital is between firms and a fixed-effects regression of k_{it} on l_{it} delivers $R^2 \approx 0.50$.

A.1 Choice of Labour and Material Inputs

Conditional on its information set, the firm solves

$$\max_{L_{it},M_{it}} \alpha K_{it}^{\beta_k} L_{it}^{\beta_l} \exp(\omega_{it}) \quad \text{s.t.} \quad M_{it} \geq \beta_m^{-1} \alpha K_{it}^{\beta_k} L_{it}^{\beta_l} \exp(\omega_{it}),$$

yielding in optimality

$$L_{it}^* = \left(\frac{\alpha\beta_l}{W_{it}}\right)^{\frac{1}{1-\beta_l}} K_{it}^{\frac{\beta_k}{1-\beta_l}} \exp\left(\frac{\omega_{it}}{1-\beta_l}\right)$$
$$M_{it}^* = \beta_m^{-1} \alpha K_{it}^{\beta_k} {L_{it}^*}^{\beta_l} \exp(\omega_{it})$$

Following ACF, we contaminate L_{it}^* with iid optimisation noise to break the functional dependence issue.

$$L_{it} = L_{it}^* \exp(\xi_{it}^l)$$

where ξ_{it}^l follows a normal distribution with mean zero and a standard deviation of 0.1. ACF chooses a standard deviation of the optimisation error of 0.37. We deliberately adopt a smaller value. A larger variance indeed strengthens identification by injecting more cross-sectional variation into the labour input and thereby mitigating—but not eliminating—the weak-instrument problem. For expositional clarity, however, a modest reduction in the variance still breaks the functional dependence at the heart of the ACF critique while keeping the simulated distributions readily interpretable.

A.2 Investment Choice and Capital Accumulation

Capital is a dynamic input. Investment is chosen at $t-1$ and the capital stock evolves as

$$K_{i,t+1} = (1-\delta)K_{it} + I_{it}, \qquad c_i(I_{it}) = \tfrac{\phi_i}{2} I_{it}^2,$$

where $\phi_i > 0$ is firm-specific; we draw ϕ_i^{-1} lognormally with cross-sectional standard deviation 0.6. Firms maximize the expected present value of profits net of adjustment costs. With the static labor choice contaminated by multiplicative optimization error,

$$L_{it} = L_{it}^* \exp(\xi_{it}^l), \qquad \xi_{it}^l \sim \mathcal{N}(0, \sigma_{\xi_l}^2) \text{ i.i.d.},$$

the (expected) marginal value of one more unit of installed capital next period is

$$\text{MVPK}_{i,t+1} = \left(\frac{\beta_k}{1-\beta_l}\right) \alpha^{\frac{1}{1-\beta_l}} W_{i,t+1}^{-\frac{\beta_l}{1-\beta_l}} \exp\left(\frac{\omega_{i,t+1}}{1-\beta_l}\right) \left[\beta_l^{\frac{\beta_l}{1-\beta_l}} e^{\frac{1}{2}\beta_l^2\sigma_{\xi_l}^2} - \beta_l^{\frac{1}{1-\beta_l}} e^{\frac{1}{2}\sigma_{\xi_l}^2}\right].$$

The term in square brackets is due to the optimization-error in labor. The Euler equation with quadratic adjustment costs yields a closed-form policy:

$$I_{it} = \frac{\beta}{\phi_i} \sum_{\tau=0}^{\infty} \big(\beta(1-\delta)\big)^{\tau}\ \mathbb{E}_t[\text{MVPK}_{i,t+1+\tau}]\,.$$

We assume $\ln W_{it}$ follows an independent covariance-stationary AR(1):

$$\ln W_{t+h} = \rho_W^h \ln W_t + \zeta_{t+h},$$

with $\zeta_{t+h} \sim \mathcal{N}\big(0, \sigma_\zeta^2 \sum_{j=0}^{h-1} \rho_W^{2j}\big)$, independent of current information. Then conditional log-normality implies, for $h = \tau + 1$,

$$\mathbb{E}_t\left[e^{\frac{\omega_{t+h}}{1-\beta_l}}\, W_{t+h}^{-\frac{\beta_l}{1-\beta_l}}\right] = \exp\left(\frac{\rho^h}{1-\beta_l}\,\omega_t - \frac{\rho_W^h \beta_l}{1-\beta_l}\,\ln W_t + \frac{1}{2}\frac{V_{\omega,h-1}}{(1-\beta_l)^2} + \frac{1}{2}\frac{\beta_l^2 V_{W,h-1}}{(1-\beta_l)^2}\right),$$

where the forecast-error variances have closed forms (for $|\rho|, |\rho_W| < 1$):

$$V_{\omega,h-1} = \sigma_\xi^2 \sum_{j=0}^{h-1} \rho^{2j} = \sigma_\xi^2\,\frac{1-\rho^{2h}}{1-\rho^2}, \qquad V_{W,h-1} = \sigma_\zeta^2 \sum_{j=0}^{h-1} \rho_W^{2j} = \sigma_\zeta^2\,\frac{1-\rho_W^{2h}}{1-\rho_W^2}.$$

Substituting the last display into the Euler sum gives a fully explicit expression for I_{it}.

$$I_{it} = \frac{\beta}{\phi_i}\,\frac{\beta_k}{1-\beta_\ell}\,\alpha^{\frac{1}{1-\beta_\ell}}\,B_\xi \sum_{\tau=0}^{\infty} \big[\beta(1-\delta)\big]^{\tau}\ \exp\left(\frac{\rho^{\tau+1}}{1-\beta_\ell}\,\omega_{it} - \frac{\rho_W^{\tau+1}\beta_\ell}{1-\beta_\ell}\,\ln W_{it} + \frac{1}{2}V_\tau\right),$$

with

$$B_\xi = \beta_\ell^{\frac{\beta_\ell}{1-\beta_\ell}}\,e^{\frac{1}{2}\beta_\ell^2\sigma_{\xi\ell}^2} - \beta_\ell^{\frac{1}{1-\beta_\ell}}\,e^{\frac{1}{2}\sigma_{\xi\ell}^2}, \quad V_\tau = \frac{V_{\omega,\tau}}{(1-\beta_\ell)^2} + \frac{\beta_\ell^2 V_{W,\tau}}{(1-\beta_\ell)^2}$$

A.3 Steady-State Initialisation

All firms start at $K_{i0} = 0$. We simulate the model forward until the cross-section of capital stocks converges to the stationary distribution implied by the policy rule; 10 consecutive periods from that steady state are retained for estimation.

A.4 Estimation

We apply the ACF and the LP estimators described in ACF Appendix A. The CUE is described in Section 3. We use moment vector

$$f_{it}(\theta) = z_{it}\left(y_{it} - \alpha - \beta_k k_{it} - \beta_l l_{it} - \rho(y_{it-1} - \alpha - \beta_k k_{it-1} - \beta_l l_{it-1})\right), \tag{A.1}$$

with instrument vector $z_{it} = (l_{it-1}, k_{it}, m_{it-1})$.

B Tables

Table B.1: Monte-Carlo estimates ($\sigma_{\xi_l} = 0.37$)

	ACF				LP				CUE			
	β_l		β_k		β_l		β_k		β_l		β_k	
Meas.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.
0.0	0.599	0.010	0.401	0.016	0.600	0.003	0.391	0.015	0.600	0.017	0.400	0.006
0.1	0.604	0.010	0.409	0.015	0.678	0.003	0.318	0.012	0.599	0.019	0.400	0.006
0.2	0.609	0.011	0.409	0.016	0.725	0.003	0.270	0.012	0.601	0.021	0.400	0.006
0.5	0.621	0.014	0.405	0.017	0.797	0.003	0.187	0.016	0.601	0.026	0.400	0.007

1000 replications. True values of β_l and β_k are 0.6 and 0.4, respectively.

Table B.2: Rejection frequencies at 5 % nominal level ($\sigma_{\xi_l} = 0.37$)

Meas.	Wald	AR	DH
0.0	0.058	0.052	0.089
0.1	0.054	0.056	0.079
0.2	0.054	0.057	0.099
0.5	0.062	0.051	0.127

Table B.3: Monte-Carlo estimates ($\sigma_{\xi_l} = 0.1$)

	ACF				LP				CUE			
	β_l		β_k		β_l		β_k		β_l		β_k	
Meas.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.	Coef.	S.D.
0.000	0.595	0.038	0.405	0.041	0.600	0.010	0.391	0.017	0.599	0.023	0.400	0.007
0.100	0.638	0.068	0.365	0.070	0.907	0.005	2.54e4	2.42e5	0.598	0.031	0.400	0.008
0.200	0.675	0.084	0.329	0.087	0.945	0.004	1.28e6	1.27e6	0.591	0.043	0.402	0.010
0.500	-7.40e8	1.09e10	7.39e8	1.09e10	0.972	0.003	4.41e5	4.00e5	0.581	0.076	0.404	0.015

1000 replications. True values of β_l and β_k are 0.6 and 0.4, respectively. Values in red indicate explosive behavior.

Table B.4: Rejection frequencies at 5 % nominal level ($\sigma_{\xi_l} = 0.10$)

Meas.	Wald	AR	DH
0.0	0.064	0.060	0.065
0.1	0.059	0.048	0.079
0.2	0.075	0.062	0.177
0.5	0.128	0.050	0.427

Table B.5: Production function estimates (Chile, ISIC 381)

	Estimates		95% CI		BS Normality[3]
Parameter	ACF	CUE	Nonrobust[1]	Robust[2]	p-value
β_k	0.064	0.047	[0.014, 0.080]	[0.003, 0.087]	0.000
β_l	0.122	0.053	[−0.090, 0.195]	[−0.083, 0.185]	0.000
β_m	0.875	0.956	[0.834, 1.078]	[0.848, 1.072]	0.000
Returns to scale	1.060	1.060	—		—

[1] Wald.
[2] Subset S-statistic Stock and Wright (2000).
[3] Shapiro–Wilk.

Table B.6: Production function estimates (US, NAICS 33)

	Estimates		95% CI		BS Normality[3]
Parameter	ACF	CUE	Nonrobust[1]	Robust[2]	p-value
β_k	0.422	0.187	[−0.660, 1.034]	[−0.153, 0.507]	0.000
β_l	0.411	0.333	[−1.135, 1.801]	$(-\infty$, 0.787]	0.000
β_m	0.237	0.219	[0.106, 0.332]	[0.052, 0.332]	0.000
Returns to scale	1.070	0.739	—		—

[1] Wald.
[2] Subset S-statistic Stock and Wright (2000).
[3] Shapiro–Wilk.

C Figures

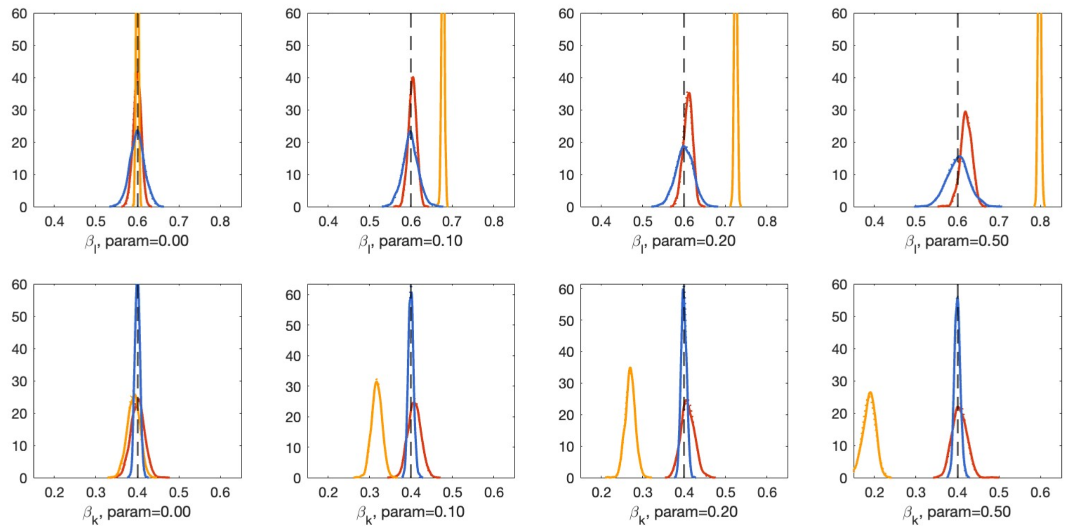

Figure C.1: 1000 replications. $\sigma_{\xi_l} = 0.37$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value. Empirical densities: solid. Normal pdf: dotted. 1000 replications.

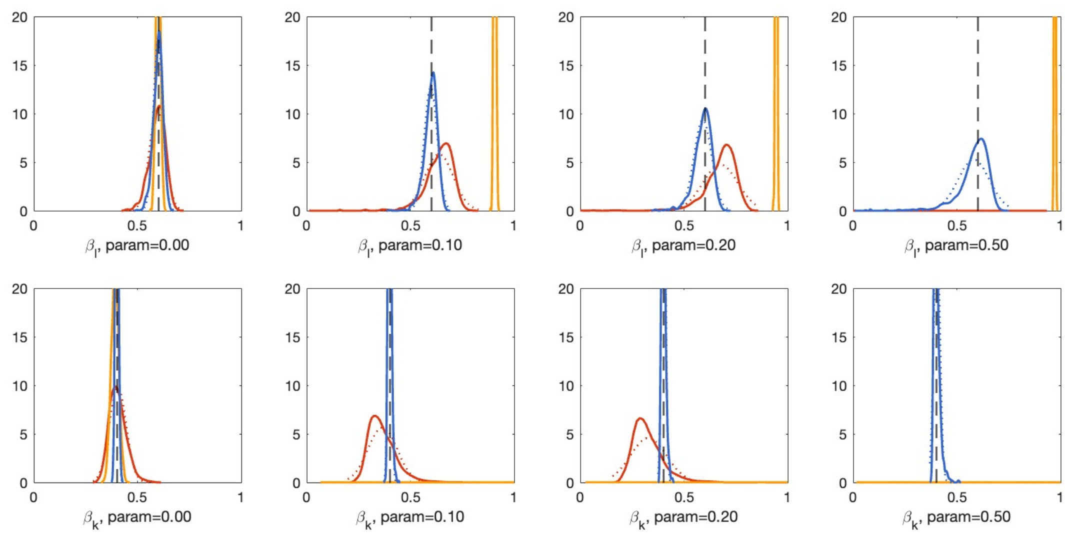

Figure C.2: 1000 replications. $\sigma_{\xi_l} = 0.10$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value. Empirical densities: solid. Normal pdf: dotted. 1000 replications.