Weak Identification Robust Methods for Production Function Estimation*

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

January 10, 2026

CLICK HERE FOR NEWEST VERSION

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. Casting control-function estimators in a GMM framework, we characterize the consequences of weak identification for estimation and inference, including non-standard asymptotic behavior and unreliable Wald inference. We then develop and implement identification-robust inference procedures for control-function estimators and provide practical guidance for diagnosing weak identification and reporting inference that remains valid when identification is weak. Finally, we illustrate how weak identification propagates to economically relevant objects constructed from production-function estimates by studying firm-level markups under the production approach.

*I am grateful for helpful comments from Isaiah Andrews, Giovanni Angelini, Stephane Bonhomme, Filippo Biondi, Giuseppe Cavaliere, Aureo de Paula, Sabien Dobbelaere, Massimo Giuliodori, Giulio Gottardo and Franc Klaassen 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, EWMES 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).

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 nonstandard asymptotic behavior and unreliable conventional inference. Second, we discuss methods for identification-robust inference. Third, we illustrate the empirical relevance of weak identification in Monte Carlo experiments and in applications to Chilean ENIA and U.S. Compustat data, and show how it propagates to derived objects such as production-approach markups.

The paper proceeds as follows. Section 2 introduces the model and characterizes the weak-identification problem in the control-function framework. Section 3 develops identification-robust inference procedures. Section 4 reports Monte Carlo experiments based on a dynamic firm model and evaluates the finite-sample behavior of conventional and robust methods. Section 5.1 applies the procedures to Chilean ENIA and U.S. Compustat data and studies the implications for production elasticities and production-approach markups. 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 and Identification

We study a standard control–function production environment in which output is generated by a general production technology, productivity follows a Markov law, with the twist that intermediate inputs provide a possibly noisy proxy for productivity.

For firm i in period t, log output satisfies

$$y_{it} = f(k_{it}, v_{it}) + \omega_{it} + \varepsilon_{it}, \tag{1}$$

where k_{it} denotes predetermined inputs (states) chosen at $t-1$ (e.g. capital) and v_{it} denotes static inputs (controls) chosen at t (e.g. labor). Let z_{it} denote the vector collecting firm i's input choices made at time t. In particular, z_{it} includes k_{it+1} and v_{it}, and may also include additional choice variables and transformations of these objects. ω_{it} is firm-specific productivity observed by the firm but not by the econometrician, and ε_{it} is an idiosyncratic shock realized after input choices that satisfies $\mathbb{E}[\varepsilon_{it}|k_{it}, v_{it}, \omega_{it}] = 0$. This formulation nests common specifications such as Cobb–Douglas, translog and CES.

Productivity follows a first-order Markov process

$$\omega_{it} = \mathbb{E}[\omega_{it}|\omega_{it-1}] + \xi_{it}, \tag{2}$$

where ξ_{it} is an innovation satisfying $\mathbb{E}[\xi_{it} \mid \omega_{it-1}] = 0$.

Firms observe ω_{it} when choosing z_{it} at time t. The shock ε_{it} is realized after these choices, while the innovation ξ_{it} is realized between $t-1$ and t. The following moment restrictions provide the basis for the control–function identification.

Assumption 2.1 (Timing and sequential exogeneity)**.** For each firm i and period t, we assume the following sequential orthogonality conditions:

$$\mathbb{E}[\varepsilon_{it} \mid z_{i\tau}] = 0, \, \tau \leq t \tag{3}$$

$$\mathbb{E}[\xi_{it} \mid z_{i\tau-1}] = 0, \, \tau \leq t \tag{4}$$

The key implication of OP/LP/ACF is that observables $p_{it} \subset z_{i\tau}$, $\tau \leq t$ can be used as proxies for ω_{it}, see ACF for details.[1] It is standard to allow p_{it} to contain contemporaneous inputs k_{it} and

[1] In the canoncial OP/LP/ACF framwork, a strong implicit assumption is made that ω_{it} is measurable with respect

v_{it}.

Assumption 2.2 (Proxy relationship)**.** There exists a function h_t such that

$$\mathbb{E}[\omega_{it} \mid p_{it}] = h_t(p_{it}). \tag{5}$$

Moreover, to exploit the proxy in a dynamic setting, we impose an additional assumption.[2]

Assumption 2.3 (Index sufficiency)**.** There exists a link function $g(\cdot)$ that satisfies

$$\mathbb{E}[\omega_{it}|p_{it-1}] = g(h_{t-1}(p_{it-1})) \tag{6}$$

We treat $g(\cdot)$ parametrically, known up to a finite-dimensional parameter $\gamma \in \mathbb{R}^{d_\gamma}$. Note that whenever ω_{it} is p_{it}-measurable, i.e. $\mathbb{E}[\omega_{it}|p_{it}] = \omega_{it}$, this condition is tautological.

Identification follows two stages. The moment conditions in Assumptions 2.1 and 2.2 imply the first stage regression, which we get plugging (5) into (1)

$$\mathbb{E}[y_{it}|p_{it}] = f(k_{it}, v_{it}) + h_t(p_{it}). \tag{7}$$

Note that in general f not identified since we allow for $(k_{it}, v_{it}) \subset p_{it}$, see ACF for discussion. Hence, we can only recuperate $\mathbb{E}[y_{it}|p_{it}]$ in this first stage. The moment conditions in Assumptions 2.1 and 2.3 imply the following second-stage regression:

$$\mathbb{E}[y_{it} \mid z_{it-1}] = \mathbb{E}[f(k_{it}, v_{it}) + g(h_{t-1}(p_{it-1}))|z_{it-1}]. \tag{8}$$

to p_{it}, i.e. satisfies the scalar-unobservable condition, which precludes any proxy error. However, arguably such a requirement is highly demanding and rarely plausible in empirical applications. It rules out any form of noise in the proxy arising from, for example, measurement or optimization errors, unexpected supply- or demand-side shocks, or heterogeneity in input prices and technologies. For a detailed discussion of proxy errors, see Hu et al. (2020). Once we allow for proxy error ω_{it} is no longer a function of observables, and the conditional distribution of ω_{it} conditional on p_{it} generally has non-degenerate variance. We can still build a control-function in terms of an object that is measurable with respect to the information available to the econometrician (see Doraszelski and Li, 2025, for a detailed discussion).

[2]This restriction is the same kind of "index sufficiency" condition emphasized in Doraszelski and Li (2025): their Theorem 1 imposes that the conditional mean of next period's productivity given the instrument set depends on observables only through the first-stage index. Under the alignment $(k_{it}, v_{it}) = z_{it}$ (i.e. the same observables are used to construct the index and to form second-stage moments), their condition reduces to Assumption 2.3.

2.2 Estimation

We follow the most common approaches and estimate the control-function model by GMM. We specify the production function as

$$f(k_{it}, v_{it}) = \alpha_f + x'_{it}\beta,$$

with $\beta \in \mathbb{R}^{d_\beta}$ the parameter of interest and where vector x_{it} collects all monomials in (k_{it}, v_{it}) up to total degree L_f (excluding the constant). Moreover let's partition x_{it} into part which is predetermined, i.e. is in z_{it-1}, call it x^p_{it}, and a part which is static, i.e. is not in z_{it-1}, call it x^s_{it}, such that $x_{it} = (x^{p\,\prime}_{it}, x^{s\,\prime}_{it})'$ The dimension of β is $d_\beta = \binom{d_k+d_v+L_f}{L_f} - 1$, where d_k, d_v are the dimensions of the primitive inputs k_{it} and v_{it}, respectively. For example, $L_f = 1$ implies the familiar Cobb–Douglas specification and $L_f = 2$ yields the translog.

We specify the law of motion as

$$g(u) = \alpha_g + b(u)'\rho,$$

with $\rho \in \mathbb{R}^{d_\rho}$. The vector $b(u)$ collects the univariate monomials in u up to degree d_ρ, $b(u) = (u, u^2, \ldots, u^{d_\rho})'$.

We approximate the control function by a polynomial sieve,

$$h_t(p_{it}) = \alpha_h + p'_{it}\lambda,$$

where p_{it} collects all monomials in the primitive proxy variables, including x_{it}, up to total degree L_h (excluding the constant). The coefficient vector is $\lambda \in \mathbb{R}^{d_\lambda}$.

The first- and second-stage restrictions (7)–(8) imply the following convenient estimation system. Since the sieve vector p_{it} nests the production-function regressors x_{it}, there exists a known selection matrix P such that $x_{it} = P\,p_{it}$. It follows that the first-stage regressors can be written as a single linear form in p_{it}:

$$x'_{it}\beta + p'_{it}\lambda = p'_{it}\gamma,$$

with $\gamma := P'\beta + \lambda$. Thus the first-stage regression identifies γ, rather than (β, λ) separately:

$$y_{it} = \alpha_1 + p'_{it}\gamma + \varepsilon_{1,it}, \tag{9}$$

where $\varepsilon_{1,it}$ is a reduced form error.

For the dynamic restriction, using $x_{it-1} = P\,p_{it-1}$ and the definition of γ yields

$$p'_{it-1}(\gamma - P'\beta) = p'_{it-1}\lambda,$$

so the dynamic index depends on (β, γ) only through λ. Substituting this index into the parametric law of motion $g(u) = \alpha_g + b(u)'\rho$ gives the second-stage equation

$$y_{it} = \alpha_2 + x'_{it}\beta + b\big(p'_{it-1}(\gamma - P'\beta)\big)'\rho + \varepsilon_{2,it} \tag{10}$$

with reduced-form error $\varepsilon_{2,it}$.

Let $\theta := (\alpha_1, \alpha_2, \beta', \gamma', \rho')' \in \Theta \subset \mathbb{R}^{d_\theta}$. We define the stacked residual vector:

$$r_{it}(\theta) := \begin{pmatrix} r_{1,it}(\theta) \\ r_{2,it}(\theta) \end{pmatrix} = \begin{pmatrix} y_{it} - \alpha_1 - p'_{it}\gamma \\ y_{it} - \alpha_2 - x'_{it}\beta - b\big(p'_{it-1}(\gamma - P'\beta)\big)'\rho \end{pmatrix},$$

Note that α_1 and α_2 absorb the constant terms in f, h_t, and g.

For the first stage, we can instrument p_{it} by itself, since $p_{it} \subset z_{it}$. In the second stage we need to instrument also for x_{it} but using lagged inputs x_{it-1}, which are included in p_{it-1}. Additionally, since the second stage residual also feature higher order polynomials in p_{it-1}, we include all monomials in p_{it-1} up to total degree d_ρ (excluding the constant), denoted as $\tilde{p}_{it-1} \in \mathbb{R}^{d_{\tilde{p}}}$, with $d_{\tilde{p}} = \binom{d_\lambda + d_\rho}{d_\rho} - 1$. Let the instrument vectors for the two equations be $w_{1,it} = \big(1,\ p'_{it}\big)' \in \mathbb{R}^{d_{w1}}$ and $w_{2,it} = \big(1,\ \tilde{p}'_{it-1}\big)' \in \mathbb{R}^{d_{w2}}$. Stack the two instrument sets in block-diagonal form,

$$W_{it} := \begin{pmatrix} w_{1,it} & 0 \\ 0 & w_{2,it} \end{pmatrix},$$

and define the stacked moment contribution

$$\phi_{it}(\theta) = W'_{it}\, r_{it}(\theta), \qquad t = 2, \ldots, T.$$

The special, but common case that g is assumed to be e linear AR(1), i.e. $d_\rho = 1$, results in a particularly straightforward GMM estimation problem. We can substitute $r_{1,it}$ into $r_{2,it}$ which

allows to directly estimate the second stage using a single equation.

$$y_{it} = \alpha_s + x_{it}'\beta + \rho\big(y_{it-1} - x_{it-1}'\beta\big) + \varepsilon_{s,it},$$

with $\varepsilon_{s,it} := \varepsilon_{2,it} - \rho\,\varepsilon_{1,it-1}$. The resulting residual function is

$$r_{s,it}(\theta) = y_{it} - \alpha_2 - x_{it}'\beta - \rho\left(y_{it-1} - x_{it-1}\beta\right)$$

This obviates the estimation of γ altogether. The second stage instrument vector $w_{2,it}$ is still valid here.

$$\phi_{it}(\theta) = w_{2,it}\, r_{s,it}(\theta), \qquad t = 2, \ldots, T.$$

The identifying restrictions are

$$\mathbb{E}[\phi_{it}(\theta_0)] = 0, \qquad t = 2, \ldots, T.$$

for some $\theta_0 \in \Theta$.

Let

$$\bar{\phi}_n(\theta) := \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=2}^{T} \phi_{it}(\theta), \qquad Q_n(\theta) := n\,\bar{\phi}_n(\theta)' W_n \bar{\phi}_n(\theta),$$

where W_n is positive definite. The joint GMM estimator is

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

We implement either two-step GMM (with $W_n = \widehat{V}(\hat{\theta}_1)^{-1}$) or the continuously updated estimator (CUE) (with $W_n = \widehat{V}(\theta)^{-1}$), where $\widehat{V}(\cdot)$ is a cluster-robust estimator of $\mathrm{Var}\big(\sqrt{n}\,\bar{\phi}_n(\theta)\big)$ based on firm-level stacked moments.

With T fixed and $n \to \infty$, we allow for arbitrary heteroskedasticity and within-firm serial dependence by estimating the moment covariance from firm-level aggregates. Let

$$\phi_i(\theta) := \frac{1}{T-1} \sum_{t=2}^{T} \phi_{it}(\theta), \qquad \bar{\phi}_n(\theta) := \frac{1}{n} \sum_{i=1}^{n} \phi_i(\theta),$$

and set

$$\widehat{V}(\theta) := \frac{1}{n}\sum_{i=1}^{n}\big(\phi_i(\theta) - \bar{\phi}_n(\theta)\big)\big(\phi_i(\theta) - \bar{\phi}_n(\theta)\big)'.$$

2.3 Misspecification

In practice, the control-function environment are at best approximations: the functional form of the production function, the law of motion, the timing restrictions, etc. Rather than interpreting the model as literally correct, it is useful to adopt the *pseudo-true parameter* perspective (see Newey and McFadden, 1994).

When the model is correctly specified, there exists θ_0 such that the moment conditions are satisfied, $E[\phi_{it}(\theta_0)] = 0$. Under misspecification, this must not be the case. Rather, they are approximations within a restricted model class. Under misspecification, extremum estimators, including GMM converge to a pseudo-true parameter: the value within the maintained model class that optimizes the population objective function. Formally, if $\hat{\theta}_n$ approximately maximizes a sample criterion $\widehat{Q}_n(\theta)$, then under standard regularity conditions it converges to

$$\theta^\star = \arg\min_{\theta\in\Theta} Q(\theta)$$

with $Q(\theta) = \mathbb{E}\big[Q_n(\theta)\big]$, which need not coincide with any structural "true" parameter when the model is misspecified. Under correct specification, standard GMM implementations that use a $\sqrt{n}$-consistent weighting matrix are first-order asymptotically equivalent (they share the same $\sqrt{n}$ limit and asymptotic variance); see the extremum equivalence results in Newey and McFadden (1994) and, for CUE, Hansen et al. (1996). Under misspecification, different choices of weighting matrix generally imply different pseudo-true limits in overidentified GMM; see Hall and Inoue (2003).

The identification analysis below is conducted conditional on the maintained moment restrictions and parametric structure. The results therefore describe how the GMM criterion behaves within this model class. To keep the exposition focused, we abstract from misspecification in the subsequent identification analysis.

2.4 Identification failure

Identification of θ_0 in the system GMM problem requires that the population moments $\mathbb{E}[\phi_{it}(\theta)]$ uniquely pin down θ in a neighborhood of θ_0. A necessary condition for *local* identification is that

the Jacobian of the population moments,

$$\mathcal{J}(\theta_0) := \mathbb{E}\left[\frac{\partial \phi_{it}(\theta)}{\partial \theta'}\right]_{\theta=\theta_0},$$

has full column rank d_θ (see e.g. Hansen, 1982; Newey and McFadden, 1994). Since the first stage (9) is a linear regression of y_{it} on p_{it}, γ is generically well identified under standard nonsingularity conditions for $\mathbb{E}[p_{it}p_{it}']$. We therefore focus on identification of the parameters that are only disciplined by the dynamic restriction in the second stage, namely (β, ρ) in the second-stage moments.

Local identification of (β, ρ) requires that the second-stage population moment map $m(\beta, \rho) := \mathbb{E}[\phi_{2,it}(\beta, \rho)]$ is locally one-to-one at (β_0, ρ_0). A necessary condition is that the Jacobian

$$\mathcal{J}_2(\theta_0) := \mathbb{E}\left[\frac{\partial \phi_{2,it}(\theta)}{\partial (\beta', \rho')}\right]_{\theta=\theta_0}$$

has full column rank $d_\beta + d_\rho$, where $\phi_{2,it}(\theta) = w_{2,it}\, r_{2,it}(\theta)$ and $\theta = (\beta', \rho')'$. This is the standard rank condition for local identification in GMM (see e.g. Hansen, 1982; Newey and McFadden, 1994). When the Jacobian is rank-deficient, the GMM criterion is locally flat in at least one direction, so that distinct values of (β, ρ) generate the same population moments and the model is not locally identified; see Mavroeidis (2005) for related identification failures in forward-looking moment conditions and Stock and Wright (2000) for the implications of weak or near-rank failure in GMM.

Working with de-meaned data, intercepts can be suppressed. Since γ is identified from the first stage, the second-stage residual satisfies

$$r_{2,it}(\theta) = y_{it} - x_{it}'\beta - b\big(p_{i,t-1}'(\gamma - P'\beta)\big)'\,\rho.$$

Let $u_{i,t-1} := p_{i,t-1}'\lambda_0$ and define

$$q_{it} := \big(x_{it}',\, \tilde{p}_{i,t-1}'\big)' \in \mathbb{R}^{d_q}, \qquad w_{2,it} := \tilde{p}_{i,t-1} \in \mathbb{R}^{d_{w2}},$$

where $\tilde{p}_{i,t-1}$ collects all nonconstant monomials in $p_{i,t-1}$ up to total degree d_ρ. By construction, $\tilde{p}_{i,t-1}$ contains $x_{i,t-1}$ and spans $(\dot{b}(u_{i,t-1})'\rho_0)\, x_{i,t-1}$, so there exists a matrix $B \in \mathbb{R}^{d_q \times d_\beta}$ such that

$$\left.\frac{\partial r_{2,it}(\theta)}{\partial \beta'}\right|_{\theta_0} = -x_{it}' + \big(\dot{b}(u_{i,t-1})'\rho_0\big)\, x_{i,t-1}' = q_{it}'B.$$

Here $\dot{b}(\cdot)$ is the derivative the polynomial basis $b(\cdot)$, we have $\dot{b}(u) = (1, 2u, \ldots, d_\rho u^{d_\rho - 1})'$, a polynomial basis of order $d_\rho - 1$. Likewise, since $b(u_{i,t-1})$ lies in the linear span of $\tilde{p}_{i,t-1}$ there exists $\Lambda \in \mathbb{R}^{d_\rho \times d_{\tilde{p}}}$ such that

$$\left.\frac{\partial r_{2,it}(\theta)}{\partial \rho'}\right|_{\theta_0} = -b(u_{i,t-1})' = \tilde{p}'_{i,t-1}\Lambda.$$

Premultiplying by $w_{2,it}$ and taking expectations yields the Jacobian decomposition

$$\mathcal{J}_2(\theta_0) = \left[\mathcal{J}_{2,\beta}(\theta_0) \,\vdots\, \mathcal{J}_{2,\rho}(\theta_0)\right] = \mathbb{E}[w_{2,it} q'_{it}]\, C,$$

where $C = \left[B \,\vdots\, \Lambda\right]$. This representation isolates the object that governs identification: the "effective" second-stage instrument variation is summarized by the cross-moment matrix $\mathbb{E}[w_{2,it} q'_{it}]$. If $\mathbb{E}[w_{2,it} q'_{it}]$ is rank-deficient, then no choice of the parametric components (B, Λ) can restore full rank in the Jacobian, so local identification must fail. Intuitively, rank failure occurs when the span of the second-stage instruments $w_{2,it}$ does not generate enough independent variation in the relevant regressors q_{it} to separately move the β- and ρ-components of the conditional mean. In control-function settings, this can arise from "functional dependence" of current inputs on lagged states and productivity (see ACF), or more generally when the reduced form induced by firm policies lies in the same low-dimensional function class as the structural components being separated (see Mavroeidis, 2005).

Proposition 2.4 (Necessary rank condition)**.** *For (β_0, ρ_0) to be locally identified, we require*

$$\operatorname{rank}\Big(\mathbb{E}[w_{2,it} q'_{it}]\Big) \;\geq\; d_\beta + d_\rho. \tag{11}$$

Proof. Local identification requires $\operatorname{rank}(\mathcal{J}_2(\theta_0)) = d_\beta + d_\rho$. Since $\mathcal{J}_2(\theta_0) = \mathbb{E}[w_{2,it} q'_{it}]\, C$ for a constant matrix C,

$$\operatorname{rank}\big(\mathcal{J}_2(\theta_0)\big) \;\leq\; \operatorname{rank}\Big(\mathbb{E}[w_{2,it} q'_{it}]\Big),$$

which implies (11). □

Condition (11) is only necessary: even when $\mathbb{E}[w_{2,it} q'_{it}]$ has sufficiently large rank, local identification may still fail if the parametric derivative matrix C is itself rank-deficient. Nevertheless, (11) provides a transparent diagnostic: identification cannot obtain unless the second-stage instruments generate at least $d_\beta + d_\rho$ linearly independent directions in the space spanned by q_{it}. In the next

subsection we give economically interpretable cases in which rank($\mathbb{E}[w_{2,it}q_{it}']$) collapses, delivering nonidentification or weak identification in the sense of Stock and Wright (2000).

2.5 Weak Identification

We show that $\hat{\beta}$ cannot be consistently estimated under weak identification. 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 Jacobian of the population moment conditions becomes arbitrarily close to losing rank. Specifically, a subset of the Jacobian's columns is local to zero so that the corresponding parameters are only weakly identified, while the remaining parameters retain standard identification.

Let $\theta = (\beta', \rho')'$, with $\beta \in \mathbb{R}^{d_\beta}$ and $\rho \in \mathbb{R}^{d_\rho}$. Define the population moments as

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

Using an exact second-order Taylor expansion, we can write

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

where $M_{\beta,n}(\rho) = \mathbb{E}\big[\partial\bar{\phi}_n(\beta, \rho)/\partial\beta'\big]$ and $M_{\rho,n}(\beta) = \mathbb{E}\big[\partial\bar{\phi}_n(\beta, \rho)/\partial\rho'\big]$. Throughout, we impose standard regularity conditions ensuring that $\mathrm{rem}_n(\beta, \rho)$ is uniformly negligible under the local sequences considered below.

Assumption 2.5 (Partial local-to-zero Jacobian)**.** Along a sequence of DGPs indexed by n,

(i) $M_{\beta,n}(\rho) = n^{-1/2}\, C(\rho)$ where $C(\rho) \in \mathbb{R}^{d_{w2} \times d_\beta}$ is continuous in ρ and bounded on Θ.

(ii) $M_{\rho,n}(\beta_0) = R$ where $R \in \mathbb{R}^{d_{w2} \times d_\rho}$.

(iii) The matrix $\big[\, C(\rho) \;\; R \,\big] \in \mathbb{R}^{d_{w2} \times (d_\beta + d_\rho)}$ has full column rank uniformly in $\rho \in \Theta_\rho$.

Define the centered empirical process on Θ as

$$\psi_n^f(\theta) \;:=\; \sqrt{n}\,\big\{\bar{\phi}_n(\theta) - m_n(\theta)\big\}.$$

Assumption 2.6 (Weak convergence)**.** The process ψ_n^f converges weakly to a mean–zero Gaussian

process ψ_0^f:

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

Assumption 2.7 (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 2.8 ($\sqrt{n}$-consistency of the strongly identified parameter)**.** *Let Assumptions 2.5–2.7 hold. Then, for any GMM minimizer $\hat{\theta} = (\hat{\beta}', \hat{\rho}')'$ of $Q_n(\theta) = n\,\bar{\phi}_n(\theta)' W_n \bar{\phi}_n(\theta)$,*

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

By Proposition 2.8, it suffices to consider local $\sqrt{n}$-perturbations of ρ when characterizing the large-sample behavior of the criterion. Let

$$\tilde{\rho} = \rho_0 + \frac{\lambda}{\sqrt{n}}, \qquad \lambda \in \mathbb{R}^{d_\rho},$$

so that $\tilde{\rho} \xrightarrow{p} \rho_0$ and $\sqrt{n}\,(\tilde{\rho} - \rho_0) = O_p(1)$.

Proposition 2.9 (Consequences of weak identification)**.** *Suppose that Assumptions 2.5–2.7 hold. Then the limiting process for the GMM objective function admits the representation*

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

Proposition 2.9 parallels Stock and Wright (2000, Theorem 1). Because $C(\rho)$ is bounded, $Q_n(\beta, \tilde{\rho})$ is stochastically bounded uniformly in β. Hence β cannot be consistently estimated: $\hat{\beta}$ is only $O_p(1)$ and has a generally non-normal limiting distribution. The behavior of the strongly identified parameter ρ is also affected. Although $\hat{\rho}$ is consistent, its limiting distribution is non-normal 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 nearly minimize 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 and asymmetric sampling behavior of GMM estimators documented in simulation studies and formalized by Stock and Wright (2000).

2.6 Example

This section specializes Section 2.1 to a parametric environment used in our Monte Carlo experiments in Section 4.

In this model, capital k_{it} is predetermined, while labor ℓ_{it} and a intermediate input m_{it} are chosen flexibly at t. Let $z_{it} := (k_{i,t+1}, \ell_{it}, m_{it})'$ denote the choice vector. Firms produce according to a Cobb–Douglas value-added technology, which does not feature intermediate inputs directly

$$y_{it} = \beta_k k_{it} + \beta_\ell \ell_{it} + \omega_{it} + \varepsilon_{it}, \tag{12}$$

where ω_{it} is observed by the firm but not by the econometrician and ε_{it} is realized after input choices and satisfies $\mathbb{E}[\varepsilon_{it} \mid \{z_{i\tau}\}_{\tau \le t}] = 0$.

Productivity follows a stationary AR(1),

$$\omega_{it} = \rho\, \omega_{i,t-1} + \xi_{it}, \tag{13}$$

with innovation ξ_{it}, which satisfies $\mathbb{E}[\xi_{it} \mid \{z_{i,\tau-1}\}_{\tau \le t}] = 0$.

Consistent with Assumption 2.2, we summarize proxy informativeness by the linear reduced form

$$\omega_{it} = \delta_k k_{it} + \delta_m m_{it} + \eta_{it}, \tag{14}$$

where η_{it} is proxy error and $\mathbb{E}[\eta_{it} \mid \{z_{i\tau}\}_{\tau \le t}] = 0$. The scalar δ_m indexes proxy strength.

Combining (12)–(13) yields the one-equation representation

$$y_{it} = \beta_k k_{it} + \beta_\ell \ell_{it} + \rho\big(y_{i,t-1} - \beta_k k_{i,t-1} - \beta_\ell \ell_{i,t-1}\big) + u_{it}, \tag{15}$$

where $u_{it} = \xi_{it} + \varepsilon_{it} - \rho\, \varepsilon_{i,t-1}$. By the timing restrictions, any function of lagged choices $\{z_{i,\tau-1}\}_{\tau \le t}$ is a valid instrument for (15). Note that $y_{i,t-1}$ is endogenous, because it contains $\varepsilon_{i,t-1}$, which also

enters u_{it}. Under the proxy relation (14), $m_{i,t-1}$ is correlated with $y_{i,t-1}$ through $\omega_{i,t-1}$ and remains orthogonal to u_{it}, so it supplies excluded instrument variation for the dynamic term.

We take as baseline instrument vector

$$w_{i,t-1} := (k_{it},\, k_{i,t-1},\, \ell_{i,t-1},\, m_{i,t-1})'.$$

This choice is parsimonious and, under the maintained AR(1) productivity law and contemporaneous proxy structure, captures the relevant reduced-form variation in $\omega_{i,t-1}$ that drives instrument relevance for $y_{i,t-1}$. Because ω_{it} is first-order Markov, deeper lags dated $t-2$ and earlier should have no predictive content for $y_{i,t-1}$ once $z_{i,t-1}$ is controlled for. Making this argument formally would require specifying the full dynamic model.

Let $\theta := (\beta_k, \beta_\ell, \rho)'$. The moment conditions can be written compactly as

$$\phi_{it}(\theta) := w_{i,t-1}\, q_{it}'\, Q(\theta), \tag{16}$$

for $t = 2, \ldots, T$, where $q_{it} := \big(k_{it},\ \ell_{it},\ k_{i,t-1},\ \ell_{i,t-1},\ y_{i,t-1}\big)'$. $Q(\theta) \in \mathbb{R}^{5\times 1}$ is the coefficient vector implied by (15).

Identification of $\theta = (\beta_k, \beta_\ell, \rho)'$ requires that θ_0 is the unique solution to the population moment condition

$$\mathbb{E}[\phi_{it}(\theta)] = 0, \qquad t = 2, \ldots, T.$$

A necessary condition for local identification is that the Jacobian

$$\mathcal{J}(\theta_0) := \mathbb{E}\Big[\frac{\partial \phi_{it}(\theta)}{\partial \theta'}\Big]_{\theta=\theta_0}$$

has full column rank $d_\theta = 3$ (e.g. Hansen, 1982).

We now isolate a configuration in which $\mathcal{J}(\theta_0)$ is rank-deficient. First, consider a *degenerate proxy*: in the linear proxy relation (14), $\delta_m = 0$, so m_{it} carries no incremental information about productivity beyond capital. In that case $m_{i,t-1}$ is irrelevant for instrumenting the endogenous component of (15), and the effective instrument space generated by $w_{i,t-1}$ collapses to functions of $(k_{it}, k_{i,t-1}, \ell_{i,t-1})$.

Second, suppose capital is *sluggish* at the sampling frequency: $k_{it} = k_{i,t-1}$ a.s. Then $(k_{it}, k_{i,t-1})$ adds only one distinct regressor, and the effective instrument space is spanned by $(k_{i,t-1}, \ell_{i,t-1})$,

which is at most two-dimensional. Consequently, the three columns of $\mathcal{J}(\theta_0)$ are linearly dependent and local identification fails.

Proposition 2.10 (Degenerate proxy and sluggish capital imply non-identification)**.** *Suppose (i)* $\delta_m = 0$ *in* (14) *and (ii)* $k_{it} = k_{i,t-1}$ *a.s. Then* $\operatorname{rank}(\mathcal{J}(\theta_0)) \leq 2$*, and hence* θ_0 *is not locally identified from the moments* $\mathbb{E}[\phi_{it}(\theta)] = 0$*.*

Proof. Under (i), the instrument vector is effectively $(k_{it}, k_{i,t-1}, \ell_{i,t-1})'$ up to measurable transformations. Under (ii), $k_{it} = k_{i,t-1}$ a.s., so the instrument space is spanned by $(k_{i,t-1}, \ell_{i,t-1})'$ and has dimension at most two. Therefore $\phi_{it}(\theta)$ takes values in a two-dimensional linear subspace, which implies that the 4×3 Jacobian $\mathcal{J}(\theta_0) = \mathbb{E}[\partial\phi_{it}(\theta)/\partial\theta']_{\theta_0}$ has rank at most two. □

The equalities in (i)–(ii) are knife-edge, but they isolate the mechanism behind weak identification in this example: when the proxy adds little incremental signal and capital varies little, the moments become nearly insensitive to distinct movements in $(\beta_k, \beta_\ell, \rho)$ and $\mathcal{J}(\theta_0)$ approaches singularity. Section 4 quantifies the resulting behavior in Monte Carlo experiments.

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^*.

$$S(\theta^*) = n f_n(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} f_n(\theta^*) \tag{17}$$

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

$$CS(\alpha) = \{\theta^* : S(\theta^*) \leq \chi^2_{d_f}(1-\alpha)\} \tag{18}$$

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 AR-GMM tests on linear transformation of the parameter vector

In many applications the objects of interest are linear combinations of the structural parameters rather than the parameters themselves. Let $\theta \in \mathbb{R}^{d_\theta}$ denote the original parameter vector and

$$\psi \;:=\; R\theta \in \mathbb{R}^{d_\psi}, \qquad d_\psi \le d_\theta,$$

be a parameter of interest defined by a known matrix R of full row rank. The remaining components of θ are treated as nuisance. Any one-to-one linear reparameterization of θ leaves the S–statistic and its limiting distribution invariant, so inference on ψ can be carried out by working directly with the original GMM objective.[3]

Write $\theta = (\psi', \eta')'$ after a suitable reparameterization, where $\eta \in \mathbb{R}^{d_\eta}$ with $d_\eta = d_\theta - d_\psi$ collects the nuisance components. To test

$$H_0 :\; \psi = \psi^*,$$

define the *restricted* CUE estimator of η as

$$\hat{\eta}(\psi^*) \;:=\; \arg\min_{\eta} \; S(\psi, \eta)\big|_{\psi=\psi^*}.$$

The corresponding linear-transformation S–statistic is then $S\big(\psi^*, \hat{\eta}(\psi^*)\big)$.

Under H_0 the limiting null distribution of $S(\psi^*)$ is stochastically dominated by a chi-squared distribution with $d_f - d_\eta$ degrees of freedom, with equality when η is well identified.[4] Using the $\chi^2_{d_f - d_\eta}(1-\alpha)$ critical value therefore yields an asymptotically size-correct test that remains valid irrespective of the strength of identification of η.

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

$$CS_\psi(\alpha) \;:=\; \Big\{\psi \in \mathbb{R}^{d_\psi} :\; S\big(\psi, \hat{\eta}(\psi)\big) \;\le\; \chi^2_{d_f - d_\eta}(1-\alpha)\Big\}.$$

[3] See, for example, Kleibergen and Mavroeidis (2009) for a detailed discussion in linear IV and GMM settings.

[4] Formally, $S(\psi^*) \xRightarrow{d} Q$, where Q is bounded above in the convex order by $\chi^2_{d_f - d_\eta}$ uniformly over parameter sequences that allow for weak identification; see Kleibergen and Mavroeidis (2009, Thm. 2).

In practice, computation proceeds by gridding over ψ, solving the restricted CUE problem for $\hat{\eta}(\psi)$ at each grid point, and retaining all values that satisfy the inequality. As in the full-parameter AR–GMM sets, the resulting confidence set may be bounded, unbounded, or disconnected in the presence of weak identification.

Subset confidence sets are obtained by taking $\psi = \theta_1$ and $\eta = \theta_2$ in the partition $\theta = (\theta_1', \theta_2')'$ with $R = \big(I_{d_{\theta_1}},\, 0\big)$. Subset AR–GMM procedures are (weakly) more powerful than projection-based methods: non-rejection by the linear-transformation test implies non-rejection by the test obtained from projecting the full-parameter AR–GMM confidence set onto the ψ-space, so projection-based sets are typically more conservative (Dufour, 1997).

3.5 Alternative identification-robust tests

The AR–GMM statistic is simple and fully identification-robust but can have limited power in overidentified designs, especially when d_f is large relative to d_θ. The weak-instrument GMM literature proposes several alternative tests that retain identification-robustness while improving power; see Kleibergen (2005), Moreira (2003), and Kleibergen and Mavroeidis (2009).

Kleibergen (2005) introduces a score-type (LM) statistic that combines the information in $f_n(\theta^*)$ with that in the Jacobian $G_n(\theta^*) = \partial f_n(\theta)/\partial\theta'$. For testing $H_0 : \theta = \theta^*$, the KLM statistic can be written as

$$K(\theta^*) \;=\; n\, \dot{f}_n(\theta^*)' \Big\{ \dot{G}_n(\theta^*)\, \hat{V}_{ff}(\theta^*)^{-1}\, \dot{G}_n(\theta^*)' \Big\}^{-1} \dot{f}_n(\theta^*),$$

where $\dot{f}_n(\theta^*)$ is a standardized score constructed from $f_n(\theta^*)$, $G_n(\theta^*)$, and $\hat{V}_{ff}(\theta^*)$.[5] Under strong identification, $K(\theta^*)$ is asymptotically equivalent to the usual LM statistic. Under weak identification, it remains bounded in the convex order by $\chi^2_{d_\theta}$, so the $\chi^2_{d_\theta}(1-\alpha)$ critical value yields an asymptotically size-correct test. In many weak-instrument designs, KLM has higher power than AR–GMM, particularly with many overidentifying restrictions.

The linear-transformation and subset framework in Section 3.4 extends directly to KLM by projecting the score with respect to ψ onto the orthogonal complement of the score with respect to η; see Kleibergen and Mavroeidis (2009). The resulting subset KLM test is identification-robust and typically more powerful than the corresponding subset AR–GMM test.

[5] See Kleibergen (2005) for the explicit construction; see also Stock et al. (2002) and Kleibergen and Mavroeidis (2009).

The conventional GMM J-statistic,

$$J(\hat{\theta}) \;=\; n\, f_n(\hat{\theta})' \hat{V}_{ff}(\hat{\theta})^{-1} f_n(\hat{\theta}),$$

is $\chi^2_{d_f - d_\theta}$ under strong identification and correct specification but is not pivotal under weak identification, since the behavior of $\hat{\theta}$ changes. Kleibergen and Mavroeidis (2009) show that J can nonetheless be combined with AR and KLM to form JKLM-type tests that retain correct size under weak identification while improving power. These statistics exploit that J is informative about model fit (overidentifying restrictions), whereas AR and KLM are robust to weak identification. Implementing such JKLM combinations in our control-function setting is conceptually straightforward but more computationally demanding, and we leave this for future work.

In linear IV models with a scalar parameter, Moreira (2003) derives the conditional likelihood ratio (CLR) test, which is exactly size-correct under weak instruments and often yields shorter confidence sets than AR or KLM. Kleibergen (2005) extends this idea to GMM, proposing a modified likelihood ratio (GMM–MLR) statistic based on a quasi-likelihood for the moments. These procedures offer attractive power properties but require simulation or bootstrap methods for critical values and are therefore more involved to implement.

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:

$$Y_{it} = \min\Big\{ \alpha\, K_{it}^{\beta_k} L_{it}^{\beta_l} \exp\big(\omega_{it}\big),\; \beta_m\, M_{it}^{*} \Big\} \exp\big(\varepsilon_{it}\big) \tag{19}$$

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 $\rho = 0.7$.

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it} \tag{20}$$

ε_{it} is independent to the current information set and mean-zero. The variance of the normally distributed innovation ξ_{it} (σ^2_ξ) and the initial value ω_{i0} ($\sigma^2_{\omega_{i0}}$) 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

$$y_{it} = \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it} \tag{21}$$

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.

$$m^*_{it} = -\log(\beta_m) + \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it}, \tag{22}$$

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

$$m_{it} = m^*_{it} + \nu_{it}, \tag{23}$$

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 5.1. 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 C.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 D.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 C.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 C.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 D.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 C.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

Our empirical analysis proceeds in two steps. We first estimate production elasticities and implied returns to scale under the ACF baseline and under CUE, using Chilean survey data (ENIA) and U.S. financial-statement data (Compustat). We then combine the estimated elasticities with input expenditure shares to construct firm-level markups.

5.1 Estimation procedure

We replicate the empirical design in Raval (2023).[6] Firms' log real revenue satisfies a Cobb–Douglas revenue production function in capital, labor, and materials, with productivity evolving as a linear AR(1). As in Raval (2023), we implement the ACF control-function estimator using a third-order polynomial approximation in (k, ℓ, m) for the proxy/state relationship. We then re-estimate the same specification with our one-step CUE estimator and report identification-robust inference alongside conventional large-sample approximations.

We follow Raval (2023) in variable construction, sample selection, and the ACF implementation. Output y_{it} is log real revenue and inputs $(k_{it}, \ell_{it}, m_{it})$ are log real capital, labor, and materials. In ENIA, materials are measured as intermediate consumption. In Compustat, materials are not observed and are proxied by COGS$-$XLR, where COGS (cost of goods sold) bundles materials, direct labor, and overhead, while XLR measures labor and related expenses with incomplete coverage. We estimate the model separately by country and industry; within each sample, the maintained specification and instrument set are held fixed across estimators, so that differences in estimates reflect the choice of estimator (ACF versus CUE) rather than changes in specification.

5.2 Markups

We follow the production approach to markup estimation rooted in the wedge logic of Hall (1988) and implemented at the firm level by De Loecker and Warzynski (2012) and subsequent work (e.g. De Loecker et al., 2020). Let $v \in \{\ell, m\}$ be a flexible input and let θ_{it}^v denote its output elasticity implied by the estimated production function. Under static cost minimization in v conditional on state inputs, no adjustment costs, and firms taking input prices as given, the first-order condition implies

$$\mu_{it} \;=\; \frac{\theta_{it}^v}{s_{it}^v}, \qquad s_{it}^v \;=\; \frac{p_{it}^v v_{it}}{P_{it} Y_{it}},$$

where s_{it}^v is the revenue share of input expenditure and μ_{it} is the price markup over marginal cost. When both labor and materials satisfy the maintained conditions, μ_{it} is invariant to the choice of v.

Raval (2023) uses this invariance as a specification check and documents systematic discrepancies between labor- and materials-based markups across datasets, emphasizing the sensitivity of production-approach markups to technology non-neutrality and measurement. De Ridder et al.

[6] Data construction details and industry definitions follow App. A of Raval (2023).

(2022) further emphasize that with financial-statement data—where prices and quantities are typically unobserved— revenue-based production function estimates need not identify the level of markups, although they can remain informative about dispersion and changes. Motivated by these insights, we report markups based on both labor and materials shares and relate their behavior to the strength of identification in the underlying elasticity estimates.

A key feature of our setting is that, in the presence of weak identification, the elasticities (and hence markups) are naturally viewed as *partially identified* objects. We therefore complement the standard "plug-in" markup series based on point estimates with uncertainty summaries that propagate identification-robust elasticity bounds into markups, both for time-series moments and for the cross-sectional distribution.

Let $(\underline{\beta}_{it}^{v}, \overline{\beta}_{it}^{v})$ denote the identification-robust confidence set for the elasticity of flexible input v for firm i at time t (obtained from the Subset-S procedure). Holding input shares fixed, this induces observation-level markup bounds

$$\underline{\mu}_{it}^{v} \;=\; \frac{\underline{\beta}_{it}^{v}}{s_{it}^{v}}, \qquad \overline{\mu}_{it}^{v} \;=\; \frac{\overline{\beta}_{it}^{v}}{s_{it}^{v}},$$

where we enforce ordering of the bounds in finite samples. For each year we summarize these bounds by their cross-sectional means, $\mathbb{E}_t[\underline{\mu}_{it}^{v}]$ and $\mathbb{E}_t[\overline{\mu}_{it}^{v}]$, and overlay the corresponding mean based on the CUE point estimates, $\mathbb{E}_t[\widehat{\mu}_{it}^{v}]$. We report changes relative to an initial base year to focus on time variation.

To characterize uncertainty in the full cross-sectional distribution of markups, we treat each firm-year markup as interval-valued, $\mu_{it}^{v} \in [\underline{\mu}_{it}^{v}, \overline{\mu}_{it}^{v}]$, and construct two complementary distributional summaries.

First, for a selected year we compute Manski-style bounds on the density using histogram binning for interval data: for each bin, the lower (upper) bound is the fraction of intervals fully contained in (intersecting) the bin, scaled by bin width. This yields a nonparametric identified set for the density that depends only on the interval information.

Second, we form an "imputation-mixture" approximation by drawing $\mu_{it}^{v,*}$ independently from a uniform distribution on $[\underline{\mu}_{it}^{v}, \overline{\mu}_{it}^{v}]$ for each observation and repeating this procedure many times. Pooling the imputed draws across repetitions yields a smooth kernel density estimate that summarizes the range of distributions consistent with the elasticity bounds under a transparent benchmark

imputation scheme. We overlay this imputation-mixture density with the kernel density of the point-estimate markups $\widehat{\mu}_{it}^{v}$ computed from CUE elasticities. The comparison illustrates how much of the cross-sectional distribution is driven by sampling and identification uncertainty in the elasticities, versus dispersion in the point estimates themselves.

5.3 Data

We use two datasets that differ in measurement quality. The Chilean data are from ENIA and are plant-level, survey-based observations for *Manufacture of food products* (ISIC 311) over 1979–1996, comprising approximately 16,000 plant–year observations. Materials are recorded as intermediate consumption and are cleanly separated from labor compensation, yielding a relatively precise proxy for productivity.

The U.S. data are from Compustat and are firm-level accounting data for *Food Manufacturing* (NAICS 31) over 1970–2005, comprising approximately 1,000 firm–year observations. Relative to ENIA, key inputs are measured with more error—most importantly, materials are constructed from accounting aggregates—so the proxy–state mapping is weaker and weak-identification concerns are correspondingly more salient in the U.S. sample.

5.4 Chile (ENIA)

Table C.5 reports results for the Chilean ENIA sample (ISIC 311). In contrast to Compustat, the estimates are stable across estimators and inference is close to standard large-sample behavior. Under ACF, elasticities are $\widehat{\beta}_k = 0.040$, $\widehat{\beta}_\ell = 0.123$, and $\widehat{\beta}_m = 0.905$, implying mildly increasing returns to scale (RTS= 1.067). The CUE estimates are similar—$\widehat{\beta}_k = 0.014$, $\widehat{\beta}_\ell = 0.110$, and $\widehat{\beta}_m = 0.947$—with RTS= 1.071. Identification-robust Subset-S confidence sets are only modestly wider than the corresponding Wald intervals and remain well behaved, indicating that the moments contain substantial information about the elasticities in this sample.

Bootstrap diagnostics likewise do not point to pronounced nonstandard behavior: Shapiro–Wilk tests fail to reject normality for each elasticity ($p = 0.896$ for β_k, $p = 0.619$ for β_ℓ, and $p = 0.493$ for β_m). Overall, these patterns are consistent with a strong proxy signal in ENIA, where materials are measured directly as intermediate consumption and are cleanly separated from labor compensation.

Figure D.3 plots the evolution of average markups in Chile (ENIA, ISIC 311), separately for labor- and materials-based markups. The solid line reports the annual mean of the plug-in markups

constructed from the CUE point estimates. The shaded band reports the corresponding year-by-year bounds obtained by propagating the identification-robust elasticity intervals into observation-level markup intervals and averaging within year.[7] The bounds are relatively tight throughout the sample and track the point series closely, indicating that the time variation in average markups is informative and not driven by weak-identification uncertainty in the underlying elasticities. This mirrors the stability of the elasticity estimates and their near-standard sampling behavior in the Chilean data.

Figure D.4 compares two estimates of the cross-sectional distribution of markups in a representative year (1996). The "imputation-mixture" density (orange) is obtained by drawing, for each firm-year, a markup uniformly from its interval $[\underline{\mu}_{it}^{v}, \overline{\mu}_{it}^{v}]$ and pooling draws across repetitions; the point-estimate density (blue) is the kernel density of plug-in CUE markups $\widehat{\mu}_{it}^{v}$. The two densities are close, and the imputation-mixture distribution is only slightly more dispersed, suggesting that interval uncertainty due to partial identification is limited in this sample. In particular, relative to the U.S. case, the distributional implications of identification-robust elasticity uncertainty are modest, consistent with a strong proxy signal in ENIA and well-behaved elasticity inference.

5.5 U.S. (Compustat)

Table C.6 reports results for the U.S. Compustat sample (NAICS 31). The estimates are considerably less stable than in the Chilean survey data. Under ACF, elasticities are $\widehat{\beta}_k = 0.329$, $\widehat{\beta}_\ell = 0.132$, and $\widehat{\beta}_m = 0.561$, implying approximately constant returns to scale (RTS= 1.022). In contrast, the CUE estimates differ sharply, with a much larger capital elasticity and sign reversals for labor and materials ($\widehat{\beta}_k = 1.181$, $\widehat{\beta}_\ell = -0.170$, $\widehat{\beta}_m = -0.364$), implying decreasing returns to scale (RTS= 0.647). This divergence between ACF and CUE is suggestive of limited information in the moments about the elasticities, making estimates sensitive to the choice of estimator and weighting.

Identification-robust inference supports this conclusion. The Subset-S confidence sets are much wider than their Wald counterparts and are highly asymmetric, covering large regions of economically implausible values (e.g. $\beta_k \in [-8.219, 1.981]$ and $\beta_\ell \in [-10.170, 1.230]$). Put differently, Wald intervals substantially understate sampling uncertainty in this setting. Bootstrap normality is rejected for each elasticity based on Shapiro–Wilk tests ($p = 0.023$ for β_k, $p = 0.015$ for β_ℓ, and $p < 0.001$ for β_m), consistent with nonstandard sampling behavior under weak identification. The Hansen J-test does not reject at the CUE estimate ($p = 0.449$), so the main empirical issue is not

[7] In each year we compute $\underline{\mu}_{it}^{v} = \underline{\beta}_{it}^{v}/s_{it}^{v}$ and $\overline{\mu}_{it}^{v} = \overline{\beta}_{it}^{v}/s_{it}^{v}$ and then take cross-sectional means. The figure reports changes relative to the first sample year.

an obvious violation of the overidentifying restrictions, but rather a weak proxy signal: with Compustat materials constructed as COGS − XLR, the proxy link to latent productivity is noisy, and the resulting moment conditions have limited leverage for pinning down the elasticities.

Figure D.5 reports the U.S. (Compustat, NAICS 31) time series of mean markups, constructed analogously to the Chilean case. The solid line is the annual mean of plug-in CUE markups $\mathbb{E}_t[\widehat{\mu}^v_{it}]$, while the shaded band plots the corresponding mean bounds $\mathbb{E}_t[\underline{\mu}^v_{it}]$ and $\mathbb{E}_t[\overline{\mu}^v_{it}]$ induced by the identification-robust elasticity intervals. In stark contrast to ENIA data, the bounds are extremely wide throughout the sample (and often extend far into economically implausible regions), so that changes in average markups are only weakly informative once elasticity uncertainty is propagated into markups. This pattern is consistent with a weak proxy signal in Compustat and with the weak-identification features already evident in the elasticity estimates and their robust confidence sets.

Figure D.6 illustrates the same issue for the cross-sectional distribution in 1996. The kernel density based on plug-in CUE markups is tightly concentrated, reflecting the cross-sectional dispersion in the point estimates. By contrast, the imputation-mixture density based on interval markups $\mu^v_{it} \in [\underline{\mu}^v_{it}, \overline{\mu}^v_{it}]$ is extremely diffuse and nearly flat over a wide range. This gap indicates that, in the U.S. sample, a substantial part of what appears as a precise distribution under plug-in markups is not supported once identification-robust elasticity uncertainty is accounted for: the data deliver limited information about the level and dispersion of markups in this setting, in line with concerns raised by De Ridder et al. (2022) for financial-statement data.

6 Conclusion

This paper revisits control-function estimation of production functions through the lens of weak identification. Even when the parameters are formally point identified, the information content of the moment conditions can be low in empirically relevant settings—most notably when the proxy variable provides only a weak signal about latent productivity. In that case the GMM criterion becomes nearly flat in economically meaningful directions, leading to unstable estimates and non-Gaussian sampling behavior, so conventional Wald-type inference can be unreliable.

To address this, we cast the control-function estimator in a GMM framework that makes the role of the moment Jacobian explicit and clarifies how weak proxy signal translates into near-rank failure.

Building on this characterization, we implement two complementary tools for applied work: (i) a bootstrap normality diagnostic that screens for weak-signal environments, and (ii) identification-robust inference based on AR/S-type statistics (and associated Subset-S confidence sets) computed from a continuously updated GMM objective with cluster-robust variance. These procedures remain valid regardless of identification strength and therefore provide a practical alternative to reporting only conventional standard errors and Wald intervals.

Empirically, the contrast between Chilean survey data (ENIA) and U.S. financial-statement data (Compustat) illustrates the stakes. In ENIA, where materials are measured as intermediate consumption, elasticity estimates are stable across estimators, bootstrap diagnostics are consistent with approximately standard behavior, and propagating identification-robust elasticity intervals into markups yields relatively tight bounds on both the time series of mean markups and the cross-sectional distribution. In Compustat, where materials must be constructed from accounting aggregates (e.g. COGS $-$ XLR), estimates are markedly less stable and identification-robust intervals become very wide; once this uncertainty is propagated into markups, both levels and distributional features become only weakly informative relative to plug-in series computed from point estimates.

For applied researchers using OP/LP/ACF-style designs—especially in settings where proxies are noisy or constructed—our recommendation is simple: report a weak-signal diagnostic together with identification-robust confidence sets, and when markups (or other nonlinear functionals) are of interest, explicitly propagate identification-robust elasticity uncertainty into those objects. This helps separate genuine economic dispersion and time variation from artifacts of weak identification and nonstandard sampling.

References

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.

De Loecker, J., Eeckhout, J., and Unger, G. (2020). The rise of market power and the macroeconomic implications. *The Quarterly Journal of Economics*, 135(2).

De Loecker, J. and Warzynski, F. (2012). Markups and firm-level export status. *American Economic Review*, 102(6):2437–2471.

De Ridder, M., Grassi, B., and Morzenti, G. (2022). Dp17532 the hitchhiker's guide to markup estimation: Assessing estimates from financial data. CEPR Discussion Paper 17532, CEPR, Paris and London.

Doraszelski, U. and Li, L. (2025). Production function estimation without invertibility: Imperfectly competitive environments and demand shocks. Working paper, 2025.

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.

Hall, A. R. and Inoue, A. (2003). The large sample behaviour of the generalized method of moments estimator in misspecified models. *Journal of Econometrics*, 114(2):361–394.

Hall, R. (1988). The relation between price and marginal cost in u.s. industry. *Journal of Political Economy*, 96(5):921–47.

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.

Hu, Y., Huang, G., and Sasaki, Y. (2020). Estimating production functions with robustness against errors in the proxy variables. *Journal of Econometrics*, 215(2):375–398.

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.

Mavroeidis, S. (2005). Identification issues in forward-looking models estimated by GMM, with an application to the phillips curve. *Journal of Money, Credit and Banking*, 37(3):421–448.

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 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.

Appendices

A Proofs of Section 2

Proof of Proposition 2.8. By Assumption 2.5 and $m_n(\beta_0, \rho_0) = 0$,

$$m_n(\beta, \rho) = n^{-1/2} C(\rho)(\beta - \beta_0) + R(\rho - \rho_0),$$

where $R = M_{\rho,n}(\beta_0)$ is constant along the sequence.

Consistency. For fixed (β, ρ),

$$\frac{1}{n} Q_n(\beta, \rho) = \bar{\phi}_n(\beta, \rho)' W_n \bar{\phi}_n(\beta, \rho) \xrightarrow{p} m_n(\beta, \rho)' W_0 m_n(\beta, \rho).$$

Because $n^{-1/2} C(\rho)(\beta - \beta_0) \to 0$, the limit reduces to

$$m_n(\beta, \rho)' W_0 m_n(\beta, \rho) \to (\rho - \rho_0)' R' W_0 R (\rho - \rho_0).$$

Since W_0 is positive definite and R has full column rank (Assumption 2.5(iii) implies $rank(R) = d_\rho$), the matrix $R' W_0 R$ is positive definite. Hence the population objective is uniquely minimized at $\rho = \rho_0$, and standard extremum arguments yield $\hat{\rho} \xrightarrow{p} \rho_0$.

$\sqrt{n}$–consistency. Write

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

where $\psi_n^f(\theta) = \sqrt{n}\{\bar{\phi}_n(\theta) - m_n(\theta)\}$. Since $\hat{\theta}$ minimizes $Q_n(\theta)$, we have $Q_n(\hat{\theta}) \le Q_n(\theta_0)$, i.e.

$$\begin{aligned} 0 &\ge Q_n(\hat{\theta}) - Q_n(\theta_0) \\ &= \big(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0) + R\sqrt{n}(\hat{\rho} - \rho_0)\big)' W_n \big(\psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0) + R\sqrt{n}(\hat{\rho} - \rho_0)\big) - \psi_n^f(\theta_0)' W_n \psi_n^f(\theta_0). \end{aligned}$$

Let $v_n := \psi_n^f(\hat{\theta}) + C(\hat{\rho})(\hat{\beta} - \beta_0)$ and $\Delta_n := R\sqrt{n}(\hat{\rho} - \rho_0)$. Then the inequality becomes

$$(v_n + \Delta_n)' W_n (v_n + \Delta_n) \;\le\; \psi_n^f(\theta_0)' W_n \psi_n^f(\theta_0).$$

Using $\lambda_{\min}(W_n)\|a\|^2 \le a' W_n a$ and the triangle inequality,

$$\lambda_{\min}(W_n)\, \|\Delta_n\|^2 \le (v_n + \Delta_n)' W_n (v_n + \Delta_n) \le \psi_n^f(\theta_0)' W_n \psi_n^f(\theta_0),$$

and therefore

$$\|\Delta_n\|^2 \le \frac{1}{\lambda_{\min}(W_n)}\, \psi_n^f(\theta_0)' W_n \psi_n^f(\theta_0).$$

Under Assumptions 2.6–2.7, $\psi_n^f(\theta_0) = O_p(1)$ and $W_n = O_p(1)$ with $\lambda_{\min}(W_n) \xrightarrow{p} \lambda_{\min}(W_0) > 0$, so the right-hand side is $O_p(1)$, implying

$$\|R\sqrt{n}(\hat{\rho} - \rho_0)\| = \|\Delta_n\| = O_p(1).$$

Finally, since R has full column rank, $\lambda_{\min}(R'R) > 0$ and

$$\|Rx\| \ge \sqrt{\lambda_{\min}(R'R)}\, \|x\| \qquad \forall x \in \mathbb{R}^{d_\rho}.$$

Applying this with $x = \sqrt{n}(\hat{\rho} - \rho_0)$ yields $\sqrt{n}(\hat{\rho} - \rho_0) = O_p(1)$. □

Proof of Proposition 2.9. Under Assumption 2.5 and the local sequence $\tilde{\rho} = \rho_0 + \lambda/\sqrt{n}$ with $\lambda \in \mathbb{R}^{d_\rho}$,

$$\sqrt{n}\, m_n(\beta, \tilde{\rho}) = C(\tilde{\rho})(\beta - \beta_0) + R\lambda.$$

Therefore,

$$\begin{aligned}\sqrt{n}\, \bar{\phi}_n(\beta, \tilde{\rho}) &= \sqrt{n}\big\{\bar{\phi}_n(\beta, \tilde{\rho}) - m_n(\beta, \tilde{\rho})\big\} + \sqrt{n}\, m_n(\beta, \tilde{\rho}) \\ &= \psi_n^f(\beta, \tilde{\rho}) + C(\tilde{\rho})(\beta - \beta_0) + R\lambda.\end{aligned}$$

By Assumption 2.6, $\psi_n^f(\beta, \tilde{\rho}) \Rightarrow \psi_0^f(\beta, \rho_0)$, and by continuity of $C(\cdot)$, $C(\tilde{\rho}) \to C(\rho_0)$. By Assumption 2.7, $W_n \xrightarrow{p} W_0$. Applying Slutsky's theorem and the continuous mapping theorem to $Q_n(\beta, \tilde{\rho}) = n\, \bar{\phi}_n(\beta, \tilde{\rho})' W_n \bar{\phi}_n(\beta, \tilde{\rho})$ yields

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

as claimed. □

B 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 $R^2 \approx 0.50$.

B.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

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

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.

B.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. This implies a strong heterogeneity, with a substantial mass of firms having very high adjustment costs. 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

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

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:

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

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 $h = \tau + 1$,

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

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

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

B.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.

B.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

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

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

C Tables

Table C.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 C.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 C.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 C.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 C.5: Elasticity estimates: Chile ENIA, ISIC 311

	Estimates		95% CI		BS normality
Parameter	ACF	CUE	Wald	Subset-S	SW p-value
β_k	0.040	0.014	[0.006, 0.022]	[0.004, 0.034]	0.896
β_ℓ	0.123	0.110	[0.087, 0.133]	[0.080, 0.150]	0.619
β_m	0.905	0.947	[0.931, 0.964]	[0.917, 0.967]	0.493
Returns to scale	1.067	1.071	[1.059, 1.084]	[1.051, 1.091]	—

Notes: Wald confidence intervals are based on the cluster-robust variance at the CUE estimate. "Subset-S" reports identification-robust confidence intervals. "SW p-value" reports the bootstrap p-value of a Shapiro–Wilk normality test for the corresponding standardized bootstrap distribution.

Table C.6: Elasticity estimates: U.S. Compustat, NAICS 31

	Estimates		95% CI		BS normality
Parameter	ACF	CUE	Wald	Subset-S	SW p-value
β_k	0.329	1.181	[−1.213, 3.576]	[−8.219, 1.981]	0.023
β_ℓ	0.132	-0.170	[−1.830, 1.489]	[−10.170, 1.230]	0.015
β_m	0.561	-0.364	[−2.069, 1.341]	[−6.964, 9.236]	< 0.001
Returns to scale	1.022	0.647	[−0.306, 1.600]	[−5.953, 7.447]	—

Hansen J-test at CUE: $J = 0.573$ (df= 1), $p = 0.449$.
Notes: "Subset-S" reports identification-robust confidence intervals. "SW p-value" reports the bootstrap p-value of a Shapiro–Wilk normality test for the corresponding estimator's bootstrap distribution.

D Figures

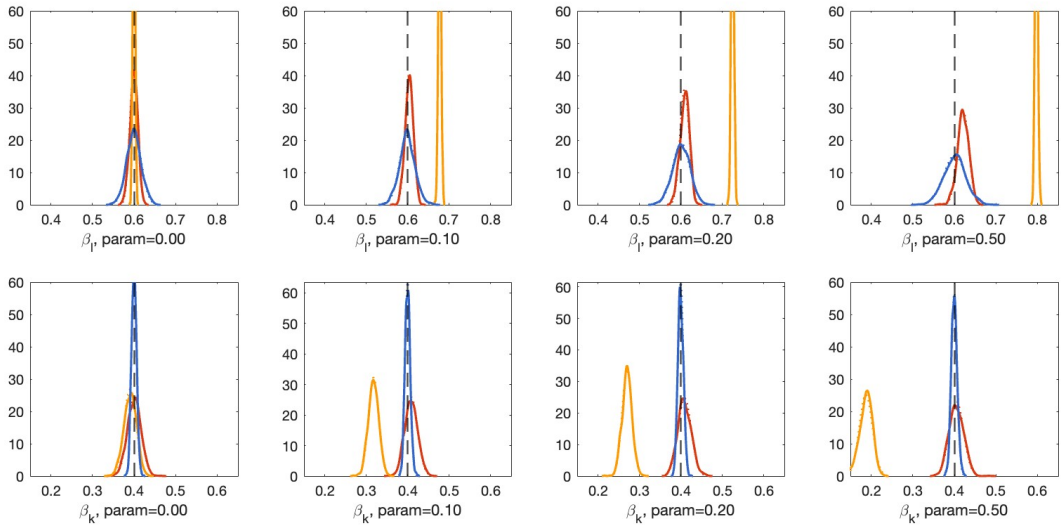


Figure D.1: 1000 replications. $\sigma_{\xi_l} = 0.37$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value.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

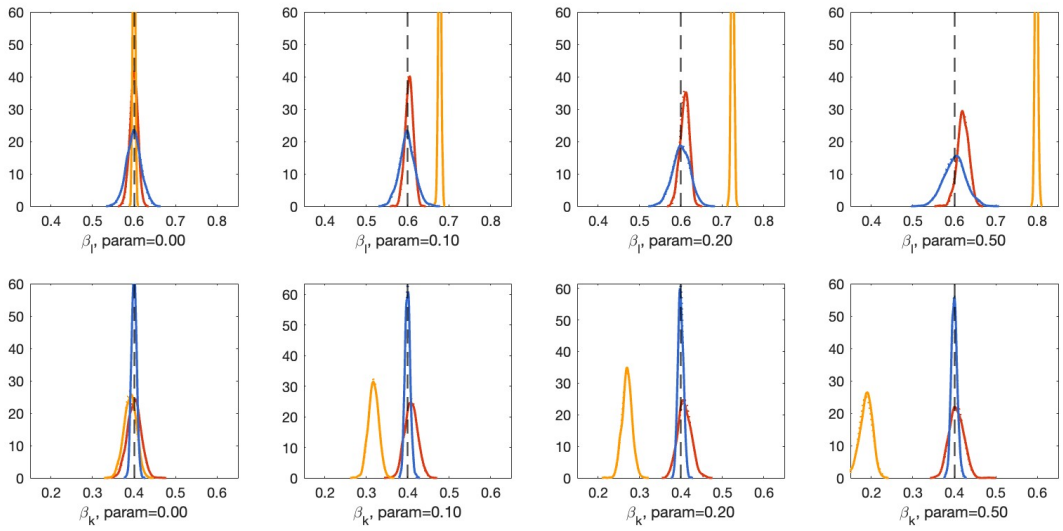
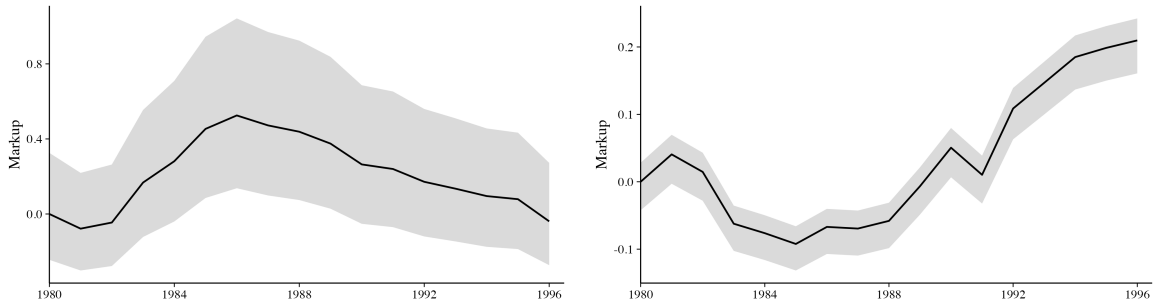


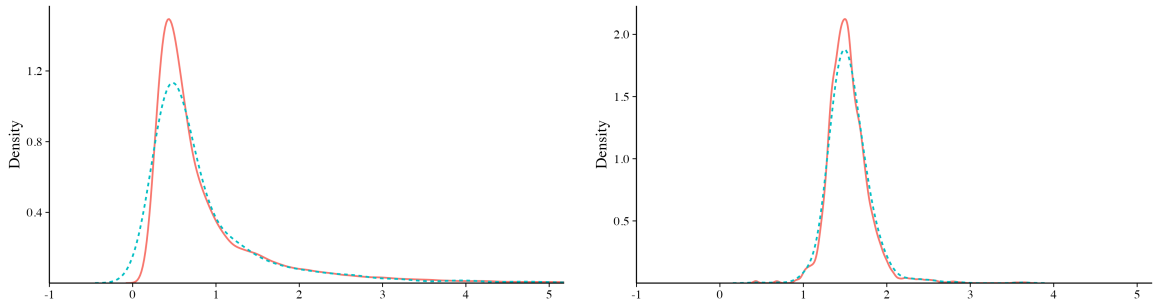
Figure D.2: 1000 replications. $\sigma_{\xi_l} = 0.10$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value.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



(a) Labor-based markups

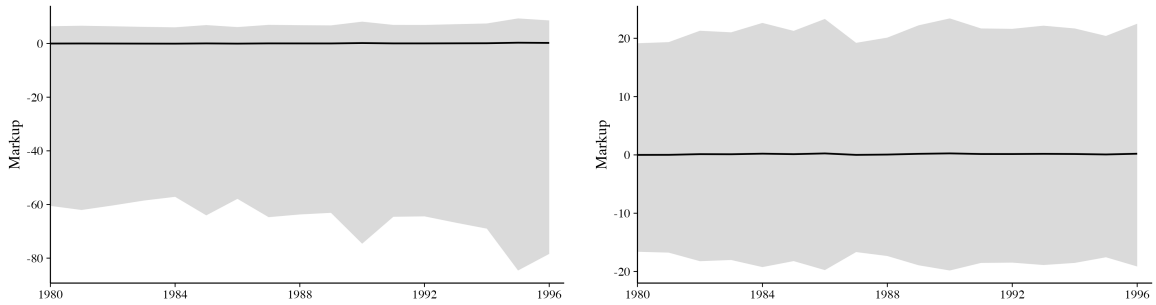
(b) Materials-based markups

Figure D.3: Chile (ENIA, ISIC 311): changes in mean markups with identification-robust bounds. Solid line: mean plug-in CUE markups. Shaded band: mean bounds induced by identification-robust elasticity intervals. Series are plotted relative to the first sample year.



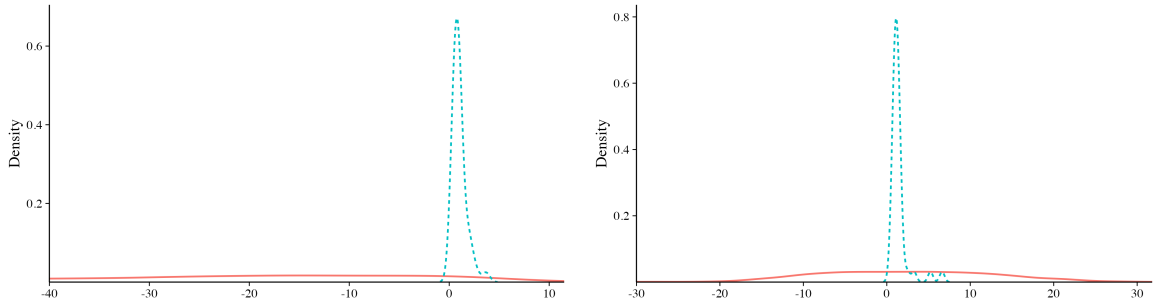
(a) Labor-based markups (b) Materials-based markups

Figure D.4: Chile (ENIA, ISIC 311): markup distributions in 1996. Dashed/red line: imputation-mixture kernel density from draws $\mu_{it}^{v,*} \sim \text{Unif}[\underline{\mu}_{it}^{v}, \overline{\mu}_{it}^{v}]$ pooled across repetitions. Solid/blue line: kernel density of plug-in CUE markups $\widehat{\mu}_{it}^{v}$.



(a) Labor-based markups (b) Materials-based markups

Figure D.5: U.S. (Compustat, NAICS 31): changes in mean markups with identification-robust bounds. Solid line: mean plug-in CUE markups. Shaded band: mean bounds induced by identification-robust elasticity intervals. Series are plotted relative to the first sample year.



(a) Labor-based markups (b) Materials-based markups

Figure D.6: U.S. (Compustat, NAICS 31): markup distributions in 1996. Dashed line: imputation-mixture kernel density from draws $\mu_{it}^{v,*} \sim \text{Unif}[\underline{\mu}_{it}^{v}, \overline{\mu}_{it}^{v}]$ pooled across repetitions. Solid line: kernel density of plug-in CUE markups $\widehat{\mu}_{it}^{v}$.