Weak Identification Robust Methods for Production Function Estimation*

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

September 19, 2025

Abstract

This paper studies weak identification in control-function estimation of production functions. We first show that identification is feasible even when the productivity proxy is noisy by formalizing a researcher-feasible single index for productivity and a forecast-sufficiency condition under which the standard moment restrictions remain valid. Identification is then characterized through an orthogonalized Jacobian that isolates the finite-dimensional parameters from the nonparametric component. When the proxy carries little signal, this Jacobian becomes nearly singular, leading to weak identification. Biased estimators with non-normal distribution are the consequences. To address this, we provide practical diagnostics of signal strength and identification-robust tests and confidence sets computed from a continuously updated GMM objective. These procedures remain valid regardless of identification strength and are straightforward to implement for applied work.

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

1 Introduction

Estimating firm-level production functions is fundamental for analyzing firm dynamics, including returns to scale, input market competitiveness, input complementarities, and technological change (Hall, 1988). Beyond microeconomic applications, firm-level production functions have an important role in empirical economic research, for instance in studies on misallocation, aggregate productivity, liquidity shocks, and market power. It is thus not surprising that the identification of production functions has received significant attention, originating with the work of Cobb and Douglas (1928) and with key developments by Marschak and Andrews (1944), Olley and Pakes (1996) (OP), Blundell and Bond (2000) and Levinsohn and Petrin (2003) (LP), and more recently by Wooldridge (2009), Ackerberg et al. (2015) (ACF) and Gandhi et al. (2020).

The identification of production functions is complicated by the simultaneity between input choices and unobserved productivity shocks, which renders inputs endogenous (Griliches and Mairesse, 1998). The standard approaches are panel data methods (Blundell and Bond, 2000), which handle endogeneity through differencing, and the *control function approach*, which uses proxy variables to account for unobserved productivity (OP, LP, ACF). In this paper, we focus on the latter.

In the control function approach, a control function that captures the relationship between the unobserved productivity shock and observable input choices is derived. This justifies using these inputs as proxies for productivity. OP first introduced this approach and chose capital investment as a proxy variable. LP extended this method by proposing intermediate inputs as alternative proxies. ACF adjusted the estimation procedure to address functional dependence between input choices. Wooldridge (2009) proposed a more efficient estimator that estimates the two sets of moments from both stages simultaneously. Gandhi et al. (2020) highlighted some limitations of the control function approach to identify production functions when the proxy variables are included in the production function and proposed an estimator without relying on proxy variables, which, on the other hand, requires stronger assumptions on the firms' behavior.

Estimation methods following the control function approach have become standard tools in empirical research and are readily available in `Stata` and `R` packages. The control function approach is extensively applied in fields such as empirical industrial organization, macroeconomics, international Trade, and labor economics. However, despite their widespread adoption, we find that the literature has largely overlooked the potential for *weak identification*. Weak identification is characterized by a case in which the model parameters are point identified but close to being unidentified. This leads to standard asymptotic theory to break down (Lewbel, 2019). In this paper we discuss weak identification issues in production function estimation.

A key practical source of weak identification is noisy proxies. In many settings the proxy (e.g., capital investment or intermediate inputs) is contaminated by measurement error, transitory shocks, or optimization frictions. These frictions weaken the link between the proxy and productivity, lowering the effective signal–to–noise ratio. As the proxy carries less reliable information, identification becomes fragile: estimates can be unstable, with a great sensitivity to starting values, and conventional standard errors misleading—closely analogous to the weak-instrument problem in IV (Staiger and Stock, 1997; Stock et al., 2002). We refer to such cases as *weak proxies*.

This paper makes several contributions. First, we show that identification is not precluded even if we relax the scalar–unobservable condition, thereby allowing for noisy proxies. If proxies are noisy, it introduces an additional latent disturbance and breaks the one–to–one mapping from the proxy to productivity. We show that the usual OP/LP/ACF moment conditions can still identify the production function parameters once we impose mild structure that delivers a one–dimensional sufficient index for predicting productivity.

Second, we uncover the potential for weak identification within the control–function approach. Framing the model in an IV language lets us leverage established diagnostics for weak identification (Bound et al., 1995; Staiger and Stock, 1997; Stock et al., 2002; Andrews and Stock, 2005; Stock and Yogo, 2005). We find that weak proxies can generate estimates with non–normal sampling behavior, undermining conventional inference.

Third, we develop a practical inference toolkit for the control-function setting with potentially weak proxies. We implement a pre-test based on bootstrapping to detect weak proxies (Angelini et al., 2024). We also conduct identification-robust inference with the S-statistic or AR-GMM statistic, which delivers valid confidence sets irrespective of identification strength (Stock and Wright, 2000). The procedures plug directly into the standard OP/LP/ACF workflow and accommodate clustered dependence.

Related literature

We contribute to several existing strands of literature. First, we add to the literature on the identification and estimation of production functions. We show that identification is feasible even when relaxing the scalar-unobservable assumption in OP/LP/ACF. Related work by Hu et al. (2020) develops a general proxy–error framework and obtains nonparametric identification under completeness–type conditions. However, our approach is closer to OP/LP/ACF: we keep the usual two–stage setup and deliver identification under structure that is straightforward to implement in standard production–function applications. Doraszelski and Lixiong (2025) make a similar point by recognizing that identification is possible under noisy proxies and adopt a slightly different but related strategy to address this. Moreover, we expand the discussion on identification in production functions by considering weak identification. While prior work has focused on deriving valid moment restrictions (see LP; ACF; Demirer 2019; Doraszelski and Jaumandreu 2013) or discuss nonidentification (Gandhi et al., 2020), much less attention has been paid to cases where the moments are nearly not uniquely satisfied—leading to weak identification. We show how noisy proxies can generate this problem.

Second, we contribute more generally to a strand of literature that has studied weak identification in various empirical applications in economics (Andrews and Guggenberger, 2019). Weak identification has been documented in a variety of non-linear econometric settings, including estimation of the New Keynesian Phillips Curve (Dufour et al., 2006; Kleibergen and Mavroeidis, 2009b; Mavroeidis, 2005; Nason and Smith, 2008), monetary-policy rules (Mavroeidis, 2010), dynamic stochastic general-equilibrium (DSGE) models (Ruge-Murcia, 2007; Canova and Sala, 2009; Iskrev, 2010; Andrews and Mikusheva, 2015; Guerron-Quintana et al., 2013), consumption capital asset pricing models (CCAPM) (Stock and Wright, 2000), and Euler equations (Yogo, 2004). These findings underscore the need for inference procedures that remain valid even when the identifying information in the data is weak.

Third, we contribute to an extensive body of work which applies these methods, by giving practical advice to empirical researchers. Some recent applications are the following. In the context of market concentration, Autor et al. (2020) and De Loecker et al. (2020) use ACF to estimate markups and analyze rising competition dynamics, while Yeh et al. (2022) and Dobbelaere et al. (2024) apply ACF to examine employer market power through price markups, wage markdowns, and collective bargaining. Related to International Trade and macroeconomics, Dobbelaere and Wiersma (2020) and Dobbelaere and Kiyota (2018) use ACF to study firm internationalization and labor market imperfections, while Caselli et al. (2021) employ ACF to estimate markups in response to global competition. Mertens (2019) apply Wooldridge's approach to assess labor market distortions, particularly those induced by trade with China. Additionally, Gopinath et al. (2017) use the Wooldridge approach to investigate capital misallocation.

Related to this is a literature that applies control function estimation of production functions to derive more complex measures. We naturally expect the uncovered identification issues to affect the reliability of these measures as well. In a seminal paper, De Loecker and Warzynski (2012) derive firm-level markups based on output elasticities, which they estimate using the control function approach, spurring a rich literature on market power, competition, and productivity across industries and countries (De Loecker et al., 2020; Karabarbounis and Neiman, 2018; Traina, 2018; Bridgman and Herrendorf, 2022; Loecker and Scott, 2017; Autor et al., 2020; Melitz and Redding, 2014; Syverson, 2019). However, Raval (2023) finds "stark differences" in these markup estimates depending on the choice of proxy variables. He attributes them to variations in firm behavior and on non-neutral productivity (Doraszelski and Jaumandreu, 2018). In light of our findings, the differences in markup estimates could also be due to weak identification, particularly if the used proxy variables differ in predictive power.

Organization of paper

This paper is structured as follows. Section 2 introduces the control function approach and shows identification weakening the scalar unobservable assumption. Section 3 discusses identification issues linked to the weak proxy problem. Section 4 presents the methods for identification robust inference. Section 5 reports Monte Carlo simulations showing the identification issues due to weak proxies and how the robust methods are still valid in face of these issues. Section 6 presents empirical results using Chilean and U.S. manufacturing data. Finally, Section 7 concludes.

2 Identification In The Control Function Approach

We briefly introduce the control function approach and discuss identification. We largely follow OP/LP/ACF. Behind the identification strategy is a discrete time model of optimizing firms $i = 1, \ldots, n$ over time periods $t = 1, \ldots, T$. Consider the following production function in logs for firm i at time t (this specification accommodates the often applied Cobb-Douglas and translog production functions):

$$y_{it} = x_{it}^{\top}\beta + \omega_{it} + \varepsilon_{it} \tag{1}$$

y is the log of output, x is the vector of the log inputs (which includes a constant to account for an intercept), namely transformations of labor and capital, ε denotes unexpected productivity shocks, and ω captures firm-specific productivity. Latin letters denote observed variables; Greek letters denote variables unobserved to the researcher. We are interested in identifying and estimating β.

At time t, firm i observes ω_{it} and then chooses $(x_{it}^f, m_{it}, x_{it+1}^p)$; ε_{it} is realized at the end of the period and y_{it} is produced. We allow for part of the input vector x_{it} to be predetermined at $t-1$, denoted at x_{it}^p and part to be flexibly determined at t, denoted as x_{it}^f, so $x_{it} = (x_{it}^p, x_{it}^f)$. In OP/LP/ACF capital is considered predetermined, whereas labor is flexible. The firm also flexibly chooses upon intermediate inputs m_{it}. We assume a "value-added specification", m_{it} do not enter (1) directly but serve as a proxy for productivity.[1] In the beginning of t period, the firm observes ω_{it} and chooses $(x_{it+1}^p, x_{it}^f, m_{it})$. At the end of period t, say at $t+b$, with $0 < b < 1$, the unexpected ε_{it} realizes and y_{it} is produced.

2.1 Model Assumptions

The model is assumed to satisfy the following assumptions. The identifying moment conditions will be based on the timing assumptions. We denote the information set of firm i at t as $\mathcal{I}_{it}$.

Assumption 1. *The information set $\mathcal{I}_{it}$, includes all current and past realizations of ω_{it}, $\{\omega_{i\tau}\}_{\tau \leq t} \subset \mathcal{I}_{it}$. The unexpected productivity shocks satisfy $\mathbb{E}[\varepsilon_{it}|\mathcal{I}_{it}] = 0$, $\forall i, t$.*

Because ω_{it} is part of $\mathcal{I}_{it}$, every input decision the firm chooses as a function of ω_{it}, is also contained in $\mathcal{I}_{it}$. Thus $x_{it+1}^p \in \mathcal{I}_{it}$.

Assumption 2. *ω follows a Markov process with law of motion g,*

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

and ξ_{it} satisfies $\mathbb{E}[\xi_{it}|\mathcal{I}_{it-1}] = 0$.

[1]For a discussion of the relationship between gross output and value-added production functions, see Bruno (1978), Basu and Fernald (1997), and Gandhi et al. (2011).

The fundamental assumption of the control function approach is that ω can be expressed in terms of observables. The canonical OP/LP/ACF setup imposes a "scalar unobservable" assumption: the firm's input choice m is pinned down by productivity ω, conditional on the predetermined input x^p. Equivalently, productivity is perfectly recoverable from the proxy variables,:

$$\omega_{it} = \mathbb{E}[\omega_{it}|m_{it}, x^p_{it}]$$

For this, the existence of a demand function for a flexibly chosen intermediate input, m is posited,

$$m_{it} = h(\omega_{it}, x^p_{it}),$$

Assuming $h(\cdot,\cdot)$ is strictly monotonic in its first argument allows to invert it, which yields the control function

$$\omega_{it} = h^{-1}(m_{it}, x^p_{it}),$$

The scalar unobservable requirement is quite demanding, as it rules out any extra error in the proxy coming from, for example, unexpected supply- or demand-side shocks, measurement or optimization error, or other possible sources of heterogeneity, such as heterogeneous input prices. For a discussion of such proxy errors, see Hu et al. (2020).

To relax the demanding scalar unobservability, we impose the following, milder assumption that allows for an unexplained part in the demand function, denoted as ν. At the same time we will impose some restrictions on ν that will allow us to achieve identification using the usual OP/LP/ACF moment conditions.

Assumption 3. *Intermediate input demand function with error*

$$m_{it} = h(\omega_{it}, x^p_{it}) + \nu_{it}, \tag{3}$$

$h(\cdot,\cdot)$ is strictly monotone in its first argument, so an inverse in that argument exists, and where ν satisfies $\mathbb{E}[\nu_{it} \mid \mathcal{I}_{it}] = 0$ and $\mathbb{E}[\nu_{it}\varepsilon_{it}] = 0$.

Assumption 3 implies that inverting $h(\omega, x^p)$ with respect to ω does not yield an expression which can completely recover ω_{it} based on (m_{it}, x_{it})

$$\omega_{it} = h^{-1}(m_{it} - \nu_{it}, x^p_{it}). \tag{4}$$

Note that setting $\nu_{it} = 0$ collapses Assumption 3 to the usual scalar-unobservable case of OP/LP/ACF. Once we allow for proxy noise ν_{it} this is no longer possible, but we can still control for the part of ω_{it} that is predictable from the researcher's data. This point is also made in Doraszelski and Lixiong (2025).

Let z_{it} collect observables (to the researcher) up to t:

$$z_{it} = \{x^p_{it+1}, x_{it}, x_{it-1}, \ldots, m_{it}, m_{it-1}, \ldots\}.$$

The moment conditions in Assumptions 1–3 are stated with respect to the firms' information set. The next assumption adapts them to the researcher's observables.

Assumption 4. *For all i, t,*

$$\mathbb{E}[\varepsilon_{it} \mid z_{it}] = 0, \qquad \mathbb{E}[\xi_{it} \mid z_{it-1}] = 0, \qquad \mathbb{E}[\nu_{it} \mid z_{it}] = 0.$$

Assumption 4 holds immediately if z_{it} is $\mathcal{I}_{it}$-measurable (i.e., $z_{it} \subset \mathcal{I}_{it}$).

Since, as laid out above, we cannot recover ω entirely, we use an *index* based on the researcher's observables. To ensure identification using in the OP/LP/ACF moment conditions is still possible, we need to to make some further assumptions on that index. As control function we use the regression function $\Psi(z) = \mathbb{E}[\omega_{it} \mid z_{it} = z]$. This allows us to define the index

$$\widetilde{\omega}_{it} := \Psi(z_{it}).$$

The proxy error $\eta_{it} := \omega_{it} - \widetilde{\omega}_{it}$ by construction satisfies $\mathbb{E}[\eta_{it} \mid z_{it}] = 0$.

For indentification, we require a forecast sufficiency property, which holds that the forecast of next period's productivity, $\mathbb{E}[\omega_{it+1} \mid z_{it}]$, depends only on $\widetilde{\omega}_{it}$. We use Γ for the conditional forecast map $u \mapsto \mathbb{E}[\omega_{it+1} \mid \tilde{\omega}_{it} = u]$; while g remains the structural law of motion in (2).

Assumption 5. *There exists a function* Γ *such that*

$$\mathbb{E}[\omega_{it+1} \mid z_{it}] \;=\; \Gamma\left(\widetilde{\omega}_{it}\right).$$

This allows us to define the forecast error as

$$\zeta_{it+1} := \omega_{it+1} - \Gamma\big(\widetilde{\omega}_{it}\big),$$

which by construction satisfies $\mathbb{E}[\zeta_{it+1} \mid z_{it}] = 0$. This moment condition will be crucial for identification.[2]

Note that Assumption 5 does not hold in general. For example, if $g(\omega) = \omega^2$, then

$$\mathbb{E}[\omega_{it+1} \mid z_{it}] = \mathbb{E}[\omega_{it}^2 \mid z_{it}] = \tilde{\omega}_{it}^2 + \mathrm{Var}(\eta_{it} \mid z_{it}),$$

so the forecast depends on the conditional variance and cannot be written as a function of $\tilde{\omega}_{it}$ alone. A leading case where Assumption 5 is true, which we will focus on later on, is when $g(\omega)$ is affine:

$$g(\omega) \;=\; \mu + \rho\,\omega, \quad \rho \in \mathbb{R}$$

For nonlinear g, Assumption 5 does not hold automatically: the conditional forecast $\mathbb{E}[\omega_{it+1} \mid z_{it}]$ may depend not only on $\tilde{\omega}_{it}$ but also on higher-order moments of the proxy error. A sufficient condition is a location–error structure: suppose $\omega_{it} = \tilde{\omega}_{it} + \eta_{it}$, where η_{it}

[2] Note that Γ can be generalized by letting it vary with (i,t), Γ_{it}, which is less restrictive on the model assumptions but requires more from the estimation.

is independent of z_{it} and its distribution is stable across (i, t). Then

$$\mathbb{E}[\omega_{it+1} \mid z_{it}] = \Gamma(\tilde{\omega}_{it}), \qquad \Gamma(u) := \mathbb{E}[g(u + \eta)].$$

As a simple illustration, if $g(\omega) = \omega^2$ and $\mathrm{Var}(\eta_{it}) = \sigma_\eta^2$, then $\Gamma(u) = u^2 + \sigma_\eta^2$.

2.2 Identification

The control–function strategy proceeds in two steps.

First stage

Substituting (1) with the definition of the control function gives

$$y_{it} = x_{it}^\top \beta + \Psi(z_{it}) + \eta_{it} + \varepsilon_{it}. \tag{5}$$

Assumption 4 implies

$$\mathbb{E}[\eta_{it} + \varepsilon_{it} \mid z_{it}] = 0. \tag{6}$$

Because x_{it} may also enter $\Psi(\cdot)$ linearly, β is not identified from (5) directly. We therefore collect terms

$$\Phi(z_{it}) := x_{it}^\top \beta + \Psi(z_{it}).$$

The corresponding regression function is

$$\Phi_0(z) := \mathbb{E}[y_{it} \mid z_{it} = z],$$

which is nonparametrically identified from the joint distribution of (y_{it}, z_{it}).[3]

[3] See Chen et al. (2014) for a general treatment of nonparametric identification in conditional moment models.

Second Stage

By Assumption 5, we can write (1) as

$$y_{it} = x_{it}^{\top}\beta + \Gamma(\widetilde{\omega}_{it-1}) + \zeta_{it} + \varepsilon_{it}. \tag{7}$$

Assumptions 4–5 imply

$$\mathbb{E}[\zeta_{it} + \varepsilon_{it} \mid z_{it-1}] = 0. \tag{8}$$

This implies the following second-stage regression function

$$\begin{aligned} \mathbb{E}[y_{it} \mid z_{it-1}] &= \mathbb{E}[x_{it}^{\top}\beta \mid z_{it-1}] \ + \ \Gamma\big(\widetilde{\omega}_{it-1}\big) \\ &= \mathbb{E}[x_{it}^{\top}\beta \mid z_{it-1}] \ + \ \Gamma\big(\Phi(z_{it-1}) - x_{it-1}^{\top}\beta\big), \end{aligned} \tag{9}$$

where in the last line we used the first-stage relation $\widetilde{\omega}_{it-1} = \Psi(z_{it-1}) = \Phi(z_{it-1}) - x_{it-1}^{\top}\beta$.

The second stage model (7)-(8) is a conditional moment model with a finite–dimensional parameter β and an infinite–dimensional nuisance function Γ. Following the semiparametric identification framework of Chen et al. (2014), identification of β proceeds by partialling out the nuisance space induced by functions of the index. In fact our model is similar to their "single IV index model". In order to show identification we will thus use their results.

Define the second–stage residual

$$\rho_{it}(\beta, \Gamma, \Phi) := y_{it} - x_{it}^{\top}\beta - \Gamma\big(\Phi(z_{it-1}) - x_{it-1}^{\top}\beta\big).$$

Assuming Φ_0 is identified from the first stage, the conditional moment restriction can be written as

$$m(\beta, \Gamma) \ := \ \mathbb{E}\big[\rho_{it}(\beta, \Gamma, \Phi_0) \mid z_{it-1}\big].$$

with $m(\beta_0, \Gamma_0) = 0$ for the true values (β_0, Γ_0).

Following Chen et al. (2014), for identification we will use the orthogonalized Jacobian matrix. For this we need the derivative of $m(\beta, \Gamma_0)$ with respect to β at β_0

$$m'_\beta = -\mathbb{E}[x_{it} \mid z_{it-1}] + \mathbb{E}\big[\Gamma'_0(\widetilde{\omega}_{it-1})\, x_{it-1} \mid z_{it-1}\big]$$

and the Gateaux derivative of $m(\beta_0, \Gamma)$ with respect to Γ at Γ_0 in direction l

$$m'_\Gamma l = -\,\mathbb{E}[l(\widetilde{\omega}_{it-1}) \mid z_{it-1}].$$

Let $\mathcal{M}$ denote the closure of the linear span of $m'_\Gamma(\Gamma_0 - \Gamma)$

$$\mathcal{M} \;:=\; \overline{\{\; \mathbb{E}[l(\widetilde{\omega}_{it-1}) \mid z_{it-1}] \;:\; \mathbb{E}[l(\widetilde{\omega}_{it-1})^2] < \infty \;\}}.$$

$\mathcal{M}$ is also called the nuisance tangent space.

For k-th unit vector in $\mathbb{R}^{d_x}$, e_k, define

$$\tau_k^* = \arg\min_{\tau \in \mathcal{M}} \; \mathbb{E}\left[(m'_\beta e_k - \tau)^\top (m'_\beta e_k - \tau)\right].$$

The orthogonalized Jacobian matrix for β is the $d_x \times d_x$ matrix Π with entries

$$\Pi_{jk} := \mathbb{E}\left[\left(m'_\beta e_j - \tau_j^*\right)^\top \left(m'_\beta e_k - \tau_k^*\right)\right], \qquad j, k = 1, \ldots, d_x.$$

For identification it is necessary that Π has full rank. Economically, the full-rank condition on Π requires that variation in (x_{it}, x_{it-1}) contains a component not spanned by functions of the index $\tilde{\omega}_{it-1}$.

Proposition 1 (Local identification of β). *Suppose Assumptions 1–5 hold and: (i)* Γ_0 *is continuously differentiable and* Γ'_0 *is Lipschitz and bounded, (ii)* $\mathbb{E}\|x_{it}\|^2 < \infty, \mathbb{E}\|x_{it-1}\|^4 < \infty$*; (iii) the matrix* Π *is nonsingular. Then* β_0 *is locally identified.*

Sketch proof. We verify the required conditions of Chen et al. (2014) to establish identification. That is, Assumptions 4–5, in order to invoke their Theorem 7.

Fréchet derivative linear and bounded. First we require that the Fréchet derivative of $m(\beta, \Gamma)$ is linear and bounded. By the expressions for m'_β and m'_Γ given above, the differential $m'(\theta - \theta_0) = m'_\beta\,(\beta - \beta_0) + m'_\Gamma(\Gamma - \Gamma_0)$ exists and is linear. Boundedness follows from conditions (i) and (ii).

Rank condition. Second, we require that Π is nonsingular, which is precisely condition (iii).

Uniform smoothness. Finally, we need uniform smoothness of m'_β. This follows from the neighborhood restriction $\mathcal{B}_r = \{\beta : \|\beta - \beta_0\| \leq r\}$ and $\mathcal{N}_\Gamma^\delta = \{\Gamma : \sup_v |\Gamma'(v) - \Gamma'_0(v)| \leq \delta\}$ for some $r, \delta \in \mathbb{R}$, and Lipschitz continuity of Γ'_0 from (i) together with finite moments from (ii).

Chen et al. (2014) Theorem 7 yields local identification of β_0.

$\square$

If we are interested in identification of Γ besides β, Chen et al. (2014) Theorem 7 give additional condition under which (β_0, Γ_0) are jointly identified (see Theorem 9). We can get to global identification by imposing global injectivity of the conditional-moment operator and a uniform rank condition (see Chen et al., 2014, Sec. 4.2)

2.3 Estimation

The two–stage identification structure carries over directly to an estimation strategy.

First stage

Following OP/LP/ACF, we estimate the first stage using a sieve estimator (Chen, 2020). We approximate the $\Phi(z_{it})$ by a P-dimensional linear sieve in z_{it}, $\Phi_P(z_{it}; \phi)$, with $\phi \in \mathbb{R}^P$

and $P \equiv P_n \to \infty$. We estimate ϕ by OLS and define the first-stage fit

$$\widehat{\Phi}_{it} := \Phi_P(z_{it}; \hat{\phi}),$$

where $\hat{\phi}$ is the OLS estimator.

Second stage

We estimate the finite–dimensional β and the infinite–dimensional Γ jointly by sieve–GMM. Let $\Gamma_J(u;\gamma)$ be a J–dimensional linear sieve in u with $\gamma \in \mathbb{R}^J$ and $J \equiv J_n \to \infty$. Define the instrument vector, $w_{it} \in \mathbb{R}^K$, as a sieve transform of the lagged observables, z_{it-1}, with $K \equiv K_n \to \infty$ and $K \geq d_x + J$. Write $\theta = (\beta, \gamma) \in \mathbb{R}^{d_x+J}$ and define the residual

$$r_{it}(\theta) := y_{it} - x_{it}^\top \beta - \Gamma_J\big(\widehat{\Phi}_{it-1} - x_{it-1}^\top \beta; \gamma\big),$$

and the moment vector

$$f_{it}(\theta) \;:=\; w_{it}\, r_{it}(\theta) \;\in\; \mathbb{R}^K.$$

From (2.4) and by the law of iterated expectations we derive the unconditional moment restriction $\mathbb{E}\big[f_{it}(\theta)\big] = 0$, which underpins our GMM estimator. Let the sample moments be

$$f_n(\theta) = \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=2}^{T} f_{it}(\theta),$$

and consider the GMM objective

$$Q_n(\theta) = n\, f_n(\theta)^\top W_n f_n(\theta),$$

with a positive definite weight matrix W_n. Let $\mathcal{B} \subset \mathbb{R}^{d_x}$ be a compact, convex set with

$\beta_0 \in \mathcal{B}$. The GMM estimator is

$$(\hat{\beta}, \hat{\gamma}) \;=\; \arg \min_{\beta \in \mathcal{B},\; \gamma \in \mathbb{R}^K} Q_n(\theta).$$

A practical choice is a two–step GMM. For this we need a consistent estimator of the covariance of the moment conditions, $V_{ff} = \mathrm{Var}(\sqrt{n} f_n(\theta))$ Let

$$f_i(\theta) = \frac{1}{T-1} \sum_{t=2}^{T} f_{it}(\theta).$$

A cluster–robust estimator, of V_{ff} that allows for arbitrary heteroskedasticity and serial correlation within each firm (see Liang and Zeger, 1986; Stock and Watson, 2008), is given by

$$\widehat{V}_{ff}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \Big(f_i(\theta) - f_n(\theta)\Big)\Big(f_i(\theta) - f_n(\theta)\Big)^{\top}.$$

Two–step GMM is asymptotically efficient among GMM estimators; see Hansen (1982) and Newey and McFadden (1994). We also consider the continuously updated estimator (CUE) of Hansen et al. (1996), which exhibits superior finite-sample behavior relative to two–step GMM. CUE minimizes $Q_n(\theta)$ using $W_n = \widehat{V}_{ff}(\theta)^{-1}$. This improves finite-sample behavior and underlies weak-identification-robust tests, as we will see in Section 4 which remain valid even when identification is weak.

A caveat is that CUE requires the variance estimator to be uniformly consistent in θ. With the above described two-stage estimator that uses a plug-in $\widehat{\Phi}$, this uniformity generally fails if we don't account for the first stage estimation error. A practical remedy is joint estimation: estimate (β, γ, ϕ) together, similar to Wooldridge (2009). Then the standard clustered $\widehat{V}_{ff}(\theta)$ is consistent uniformly on compact Θ, making CUE and the associated S-test well behaved.

2.4 Affine Process

A leading special case of the model assumes

$$g(u) = \alpha + \rho u, \tag{10}$$

with $\alpha, \rho \in \mathbb{R}$. Because $\mathbb{E}[\eta_{it-1} + \varepsilon_{it-1} \mid z_{it-1}] = 0$, We can substitute y_{it-1} in place of $\Phi(z_{it-1})$ in and use the moment condition

$$\mathbb{E}[\varepsilon_{it} + \rho\varepsilon_{it-1} - \rho\eta_{it-1} \mid z_{it-1}] = 0.$$

We get the regression function

$$\mathbb{E}[y_{it} \mid z_{it-1}] = \mathbb{E}[x_{it}^\top \beta \mid z_{it-1}] + \alpha + \rho(y_{it-1} - x_{it-1}^\top \beta) \tag{11}$$

Note that α is absorbed by the intercept, so we will henceforth disregard it.

Denote $\theta = (\beta^\top, \rho)^\top$. One can show that given Assumptions 1-5, the moment vector

$$f_{it}^o(\theta) = z_{it-1}\left(y_{it} - x_{it}^\top \beta - \rho\left(y_{it-1} - x_{it-1}^\top \beta\right)\right) \tag{12}$$

satisfies $\mathbb{E}\left[f_{it}^o(\theta)\right] = 0$.

To consistently estimate $V_{f^o f^o}(\theta) = \mathrm{Var}(\sqrt{n} f_n^o(\theta))$, while allowing for arbitrary heteroskedasticity and serial correlation within each firm, we use a cluster–robust estimator over firms. Define the time–aggregated moment for firm i and its sample average:

$$f_i^o(\theta) := \frac{1}{T-1}\sum_{t=2}^{T} f_{it}^o(\theta), \qquad f_n^o(\theta) := \frac{1}{n}\sum_{i=1}^{n} f_i^o(\theta).$$

Then the clustered variance estimator is

$$\hat{V}_{f^o f^o}(\theta) = \frac{1}{n} \sum_{i=1}^{n} \big(f_i^o(\theta) - f_n^o(\theta)\big)\big(f_i^o(\theta) - f_n^o(\theta)\big)^\top. \tag{13}$$

3 Identification Issues

We analyze identification strength in the affine-process case described above (see Section 2.4). This specification is widely used in practice (e.g., ACF), delivers closed-form moments, and makes the rank condition—and thus the sources of weak identification—transparent. It also maps our setting to a familiar linear IV structure, letting us connect proxy strength to instrument strength in a straightforward way. The conclusions can still be generalized to processes other than the affine but for illustrative purposes we stick to this simpler model.

Let $m^o(\theta) := \mathbb{E}[f_{it}^o(\theta)]$ with $\theta = (\beta^\top, \rho)^\top \in \mathbb{R}^{d_x+1}$. To attain identification we require that the Jacobian matrix evaluated at the true values has full rank. In the affine process case, the Jacobian is given by

$$\begin{aligned}\mathcal{J}(\theta) &= (\mathcal{J}_\beta(\theta) \vdots \mathcal{J}_\rho(\theta)) \\ &= -\Big(\mathbb{E}\big[z_{it-1}(x_{it}^\top - \rho\, x_{it-1}^\top)\big] \vdots \mathbb{E}\big[z_{it-1}(\Phi_{it-1} - x_{it-1}^\top \beta)\big]\Big)\end{aligned}$$

where we used Φ_{it-1} as shorthand for $\Phi(z_{it-1})$. For a unique solution in a sufficiently small neighborhood of θ_0, we need that the Jacobian evaluated at θ_0 has full column rank. This condition is necessary for local identification, and hence also for global identification (Rothenberg, 1971; Newey and McFadden, 1994).

Note that this rank condition is the this is the finite–dimensional analogue of the rank condition in Proposition 1. In the affine case the nuisance space is one–dimensional, spanned by $\mathcal{J}_\rho(\theta_0)$, so identification holds iff $\mathcal{J}_\beta(\theta_0) \notin \text{span}(\mathcal{J}_\rho(\theta_0))$; equivalently, $\mathcal{J}(\theta_0)$ has full column rank.

Identification Failure

Suppose, contrary to the rank condition, that $\mathcal{J}(\theta_0)$ is rank reduced. Then there exists a non-zero vector $v = (v_\beta^\top, v_\rho)^\top \in \mathbb{R}^{d_x+1}$ such that $\mathcal{J}(\theta_0)\, v = 0$. In that case we can construct a continuum of solutions $\theta_t := \theta_0 + t\, v$ for any $t \in \mathbb{R}$.

We demonstrate that identification fails in a stylised case where the productivity index is a linear function of collinear regressors and the proxy variable contributes no independent variation. Moreover, we assume all inputs are predetermined, so that $x_{it} \subseteq z_{it-1}$, which is most favorable for identification, since it maximizes the number of valid instruments available in the second stage.

Theorem 1 (Proxy relevance and Jacobian rank)**.** *Consider the affine process* (10) *and suppose:*

(i) Predetermined inputs: $x_{it} \subseteq z_{it-1}$.

(ii) Perfect collinearity: $x_{it-1} = c\, x_{it}$ *for some* $c \in \mathbb{R}$.

(iii) Linear control:

$$\widetilde{\omega}_{it-1} = \Psi(z_{it-1}) = x_{it-1}^\top \delta_x + \delta_m\, m_{it-1}.$$

Let $\mathcal{J}(\theta_0) = (\mathcal{J}_\beta(\theta_0) \vdots \mathcal{J}_\rho(\theta_0))$*, have blocks*

$$\mathcal{J}_\beta(\theta_0) = \mathbb{E}\big[z_{it-1}\,(x_{it} - \rho_0 x_{it-1})^\top\big]\,, \qquad \mathcal{J}_\rho(\theta_0) = \mathbb{E}[z_{it-1}\,\widetilde{\omega}_{it-1}]\,.$$

Then:

(a) If $\delta_m = 0$*, the Jacobian is rank–deficient.*

(b) If $\delta_m \neq 0$ *and the matrix* $\big(\mathbb{E}[z_{it-1}x_{it-1}^\top] \vdots \mathbb{E}[z_{it-1}m_{it-1}]\big)$ *has full column rank, then* $\mathcal{J}(\theta_0)$ *has full column rank.*

Proof. (a) With $\delta_m = 0$ the index reduces to $\widetilde{\omega}_{it-1} = x_{it-1}^{\top}\delta_x = c\, x_{it}^{\top}\delta_x$. Hence $\mathcal{J}_\beta(\theta_0) = \big(c - \rho_0\big)\, \mathcal{J}_\rho(\theta_0)$, making it rank reduced. A non-zero vector $v = (\delta_x^{\top},\, c - \rho_0)^{\top}$ lies in the nullspace of $\mathcal{J}(\theta_0)v = 0$.

(b) Suppose a vector $v = (v_\beta^{\top}, v_\rho)^{\top} \neq 0$ lies in the null-space, i.e. $\mathcal{J}(\theta_0)v = 0$. But then there exists a scalar $k \in \mathbb{R}$ such that $\delta_m \mathbb{E}[z_{it-1}m_{it-1}] = k\, E[z_{it-1}x_{it-1}^{\top}]\delta_x$ contrary to the assumption. $\square$

Theorem 1 shows that identification requires $\mathcal{J}_\rho(\theta_0)$ to add variation outside the space spanned by $\mathbb{E}[z_{it-1}x_{it}]$, which occurs when the m_{it-1} contributes independent predictive content for productivity conditional on x_{it-1}. If m_{it-1} does not add independent variation content beyond x_{it-1}, then rank condition breaks down.

Intuitively, the index $\widetilde{\omega}_{it-1}$ is needed to identify ρ, which is one-dimensional. If the index is solely a function of x_{it-1}, which is the case if $\delta = 0$, and it are is with x_{it}, then there is effectively only one valid instrument, which is insufficient for identifying the parameter vector β.

Weak Identification

While the conditions in Theorem 1—collinear regressors and a productivity index lacking independent variation—may appear stylized, they reflect issues that are commonly encountered in empirical practice. In particular, production inputs are often highly persistent over time, which induces strong correlation between current and lagged values, making x_{it} and x_{it-1} nearly collinear in short panels or environments with adjustment frictions.

At the same time, commonly used proxies such as investment or material expenditures often provide limited independent variation in productivity once input choices are controlled for. That is, the intermediate input m_{it-1} may be functionally dependent on x_{it}, especially when arising from first-order conditions or demand systems, so that the productivity index $\widetilde{\omega}_{it-1}$ is weakly correlated with m_{it-1} conditional on the inputs. We call this weak proxies.

This combination of persistent inputs and weak proxy relevance leads to a setting of weak identification, where the moment conditions do not carry sufficient information to consistently estimate β. This leads to nonstandard behavior of GMM estimators (see Stock and Wright (2000)). Consequently, point estimates may be biased or imprecise, and Wald-type statistics do not have standard asymptotic distributions, confidence intervals may fail to have correct coverage. Therefore, even in the absence of exact collinearity or complete lack of proxy relevance, weak identification may generate substantial bias and inference problems.

So, the condition on the proxies to be explanatory, i.e. $\delta_m \neq 0$, is not sufficient for reliable estimation and inference because of the problem of weak identification. This highlights the practical relevance of the rank condition and its sensitivity to the structure of inputs and proxies.

To formalize the weak identification case, we adapt the framework of Stock and Wright (2000). The key idea is that when the Jacobian is nearly rank-deficient, β no longer converges at the $\sqrt{n}$ rate. We model the near–rank deficiency as a Pitman drift, where the relevant Jacobian component shrinks toward zero at rate $1/\sqrt{n}$.

Theorem 2 (Weak-Identification Asymptotics). *Let*

$$\hat{\theta}_n \;=\; \arg\min_{\theta\in\Theta} Q_n(\theta),$$

$Q_n(\theta) = n f_n(\theta)^\top W_n\, f_n(\theta)$ *for a compact* $\Theta \subset \mathbb{R}^{d_x+d_\rho}$ *and with a weight matrix* $W_n \xrightarrow{p} W$, *where* W *is symmetric and positive definite. Partition* $\theta = (\beta^\top, \rho)^\top$, *and assume we have a root-n consistent estimator for* ρ_0, $\hat{\rho}_n$.

(i) Joint CLT:

$$\sqrt{n}\begin{pmatrix} f_n(\theta_0) \\ \hat{\rho}_n - \rho_0 \end{pmatrix} \xrightarrow{d} \begin{pmatrix} \psi_f \\ \psi_\rho \end{pmatrix} = \psi,$$

where ψ *is mean–zero Gaussian with* $Var(\psi) = \Omega$ *positive definite.*

(ii) Nearly singular Jacobian: $J_\beta(\theta_0) = \Pi_n = C/\sqrt{n}$ *for a non–singular* $C \in \mathbb{R}^{d_z \times d_x}$, *and the columns of* C *are linearly independent of* $J_\rho(\theta_0) = r \in \mathbb{R}^{d_z}$.

Then the sample moments have the following limiting representation

$$\sqrt{n}\, f_n(\beta, \hat{\rho}_n) \ \xrightarrow{d}\ \psi_f + C(\beta - \beta_0) + r\, \psi_\rho.$$

Sketch. By twice continuous differentiability, a first–order expansion of $f_n(\theta)$ around θ_0 gives

$$\sqrt{n}\, f_n(\beta, \rho) = \sqrt{n}\, f_n(\theta_0) + C(\beta - \beta_0) + r\, \sqrt{n}\, (\rho - \rho_0) + R_n(\theta).$$

where $R_n(\theta)$ collects second–order terms. By (i), $\sqrt{n}\, f_n(\theta_0) \xrightarrow{d} \psi_f$, and $\sqrt{n}(\hat{\rho}_n - \rho_0) \xrightarrow{d} \psi_\rho$. Moreover, the remainder satisfies $R_n(\beta, \hat{\rho}_n) = o_p(1)$, since $\sqrt{n}(\hat{\rho}_n - \rho_0)^2 = O_p(n^{-1/2})$. Slutsky's theorem yields

$$\sqrt{n}\, f_n(\beta, \hat{\rho}_n) \ \xrightarrow{d}\ \psi_f + C(\beta - \beta_0) + r\, \psi_\rho,$$

as claimed. □

Theorem 2 implies that the objective function is $O_p(1)$ for any fixed β so it cannot converge at rate $n^{-1/2}$. Even if $\hat{\rho}_n$ is root-n consistent, the term $C(\beta - \beta_0)$ enters at the same $O_p(1)$ scale. Hence, for any estimator $\hat{\beta}_n$ we have $\hat{\beta}_n - \beta_0 \ = \ O_p(1)$, This breakdown of standard $\sqrt{n}$–consistency for β is the consequence of weak identification.

This setting is similar to Stock and Wright (2000). They argue that the limiting distribution for θ_n is non-standard, even for root-n consistent component $\hat{\rho}_n$. As a consequence Wald-type inference becomes invalid. Identification-robust procedures are then required for valid inference.

Concentration parameter

The weak–proxy problem in our setting closely parallels the weak–instrument problem in linear IV. In IV, instrument strength is summarized by the *concentration parameter* (Stock et al., 2002), which captures how much signal the instruments provide in the reduced form, relative to noise. We can apply it to quantify identification strength in our model by considering a linearized special case.

Consider the linear IV regression model with a single endogenous regressor x, controls w and instruments z:

$$y_i = x_i\beta + w_i'\gamma + u_i, \tag{14}$$

$$x_i = z_i'\pi + w_i'\delta + v_i, \tag{15}$$

for $i = 1, \ldots, n$. The errors u_i and v_i are assumed to be i.i.d. normal with zero mean and covariance matrix $\Sigma = (\sigma_u^2, \sigma_{uv}; \sigma_{uv}, \sigma_v^2)$.

Let $z_i^{\perp}$ be the residual from projecting z_i on w_i. The concentration parameter is

$$\mu^2 \;=\; \frac{\pi^\top \, \mathbb{E}[z_i^{\perp} {z_i^{\perp}}^\top] \, \pi}{\sigma_v^2},$$

so larger μ^2 means stronger reduced-form signal in (14). When μ^2 is small (local to zero), instruments are weak; 2SLS is biased toward OLS and Wald tests are size-distorted (Staiger and Stock, 1997; Stock et al., 2002).

The one–step affine specification (cf. Section 2.4) can be written as

$$y_{it} \;=\; \pi_1 x_{it} \;+\; \pi_2 y_{it-1} \;+\; \pi_3 x_{it-1} \;+\; u_{it}, \tag{16}$$

where y_{it-1} is endogenous. Using the linear control function with proxy error, and scalar m

$$\omega_{it} \;=\; \delta_0 \;+\; \delta_m m_{it} \;+\; x_{it}^\top \delta_x \;+\; \eta_{it},$$

the reduced form for the endogenous regressor is

$$y_{it-1} \;=\; x_{it-1}^\top(\beta + \delta_x) \;+\; \delta_m \, m_{it-1} \;+\; \varepsilon_{it-1} \;+\; \eta_{it-1}, \tag{17}$$

so m_{it-1} plays the role of the excluded instrument for y_{it-1} (conditional on x_{it-1}).

Let $m_{it-1}^{\perp}$ denote the residual from projecting m_{it-1} on x_{it-1}. The analogue of the IV concentration parameter at cross-section t is then

$$\mu_t^2 \;=\; \frac{n\,\delta_m^2\,\mathrm{Var}(m_{it-1}^{\perp})}{\mathrm{Var}(\varepsilon_{it-1}+\eta_{it-1})} \;=\; \frac{n\,\delta_m^2\,\mathrm{Var}(m_{it-1}\mid x_{it-1})}{\mathrm{Var}(\varepsilon_{it-1}+\eta_{it-1})}. \tag{18}$$

Suppose the intermediate-input demand is $m_{it} = \gamma_\omega\,\omega_{it} + x_{it}^{\top}\lambda + \nu_{it}$, with $(\omega_{it}, x_{it}, \nu_{it})$ jointly normal, $\sigma^2_{\omega|x} := \mathrm{Var}(\omega_{it} \mid x_{it})$, and $\sigma^2_\nu := \mathrm{Var}(\nu_{it})$. Then the concentration parameter implied by (17) can be written as

$$\mu_t^2 \;=\; \frac{n\,\gamma_\omega^2\,\sigma^4_{\omega|x}\,\big[(x_{it-1}^{\top}\lambda)^2 + \gamma_\omega^2\sigma^2_{\omega|x} + \sigma^2_\nu\big]}{\sigma^2_\varepsilon\big(\gamma_\omega^2\sigma^2_{\omega|x} + \sigma^2_\nu\big)^2 \;+\; \sigma^2_{\omega|x}\,\sigma^2_\nu\big(\gamma_\omega^2\sigma^2_{\omega|x} + \sigma^2_\nu\big)}.$$

The concentration parameter μ_t^2 is strictly decreasing in σ^2_ν. Intuitively, as σ^2_ν increases, the intermediate input m_{it} becomes a noisier proxy for productivity ω_{it}, weakening the first-stage relationship and thus reducing identification strength. This mirrors the well-known result in linear IV models where greater measurement error leads to weak instruments. We show in Section 5 via Monte Carlo how greater measurement error also leads to weak identification in the nonlinear model.

4 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.

4.1 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).

4.2 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^*.

$$S(\theta^*) = n f_n(\theta^*)' \hat{V}_{ff}(\theta^*)^{-1} f_n(\theta^*) \tag{19}$$

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.

4.3 Robust Confidence Sets

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

$$CS(\alpha) = \{\theta^* : S(\theta^*) \leq \chi^2_{d_f}(\alpha)\} \tag{20}$$

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.

4.4 Subset AR-GMM test

Let $\theta = (\theta_1^\top, \theta_2^\top)^\top$ 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, 2009a, 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).

5 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).

5.1 Setting

Output is produced with a Leontief production function in materials:

$$Y_{it} = \min\Big\{\alpha\, K_{it}^{\beta_k} L_{it}^{\beta_l} \exp\big(\omega_{it}\big),\ \beta_m\, M_{it}^{*}\Big\}\ \exp\big(\varepsilon_{it}\big) \tag{21}$$

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 $\rho = 0.7$.

$$\omega_{it} = \rho\omega_{it-1} + \xi_{it} \tag{22}$$

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

$$y_{it} = \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it} + \varepsilon_{it} \tag{23}$$

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.

$$m_{it}^* = -\log(\beta_m) + \alpha + \beta_k k_{it} + \beta_l l_{it} + \omega_{it}, \tag{24}$$

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

$$m_{it} = m_{it}^* + \nu_{it}, \tag{25}$$

where ν_{it} is normal with zero mean and standard deviation σ_ν. We will generate data for different values of σ_ν. σ_ν adds noise to the control function. For high values 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}.

5.2 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 6. We report the conventional (non-robust) Wald test and the identification-robust AR-GMM test from Section 4 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 4, which assesses joint normality of the estimators using Shapiro–Wilk–based diagnostics.

Table 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 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 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 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 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 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.

6 Empirical Analysis

We replicate the empirical exercise in Raval (2023).[4] Specifically, we estimate a Cobb–Douglas revenue production function with capital, labor, and materials,

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

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

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.

6.1 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 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).

6.2 Chile (ENIA)

Table 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.

6.3 U.S. (Compustat)

Table 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.

7 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.

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. and Guggenberger, P. (2019). Identification- and singularity-robust inference for moment condition models. *Quantitative Economics*, 10:1703–1746. This paper introduces a new identification- and singularity-robust conditional inference method.

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

Andrews, I. and Mikusheva, A. (2015). Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 6(1):123–152.

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.

Autor, D., Dorn, D., Katz, L. F., Patterson, C., and Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. *The Quarterly Journal of Economics*, 135(2):645–709.

Basu, S. and Fernald, J. G. (1997). Returns to scale in U.S. production: Estimates and implications. *Journal of Political Economy*, 105(2):249–283.

Blundell, R. and Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3):321–340.

Bound, J., Jaeger, D. A., and Baker, R. M. (1995). Problems with instrumental variables estimation when the correlation between the instruments and the endogeneous explanatory variable is weak. *Journal of the American Statistical Association*, 90(430):443–450.

Bridgman, B. and Herrendorf, B. (2022). The economics of factor shares. *Review of Economic Dynamics*, 44:100–129.

Bruno, M. (1978). Duality, intermediate inputs, and value-added. In Fuss, M. and McFadden, D., editors, *Production Economics: A Dual Approach to Theory and Applications*, volume 2, pages 3–16. North-Holland, Amsterdam.

Canova, F. and Sala, L. (2009). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics*, 56(4):431–449.

Caselli, M., Nesta, L., and Schiavo, S. (2021). Imports and labour market imperfections: Firm-level evidence from france. *European Economic Review*, 131:103632.

Chen, X. (2020). Semiparametric methods and theory. *Annual Review of Economics*, 12:167–196.

Chen, X., Chernozhukov, V., Lee, S., and Newey, W. K. (2014). Local identification of nonparametric and semiparametric models. *Econometrica*, 82(2):785–809.

Cobb, C. W. and Douglas, P. H. (1928). A theory of production. *The American Economic Review*, 18(1):139–165.

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.

Demirer, M. (2019). Production function estimation with imperfect proxies. Preliminary draft.

Dobbelaere, S., Hirsch, B., Mueller, S., and Neuschaeffer, G. (2024). Organized labor, labor market imperfections, and employer wage premia. *ILR Review*, 77(3):396–427.

Dobbelaere, S. and Kiyota, K. (2018). Labor market imperfections, markups and productivity in multinationals and exporters. *Labour Economics*, 53:198–212. European Association of Labour Economists 29th annual conference, St.Gallen, Switzerland, 21-23 September 2017.

Dobbelaere, S. and Wiersma, Q. (2020). The impact of trade liberalization on firms' product and labor market power. Technical report, IZA - Institute of Labor Economics.

Doraszelski, U. and Jaumandreu, J. (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies*, 80(4):1338–1383.

Doraszelski, U. and Jaumandreu, J. (2018). Measuring the bias of technological change. *Journal of Political Economy*, 126(3):1027–1084.

Doraszelski, U. and Lixiong, L. (2025). Production function estimation without invertibility: Imperfectly competitive environments and demand shocks. NBER Working Paper 33939, National Bureau of Economic Research, Cambridge, MA.

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

Dufour, J.-M., Khalaf, L., and Kichian, M. (2006). Inflation dynamics and the new keynesian phillips curve: An identification-robust econometric analysis. *Journal of Economic*

Dynamics and Control, 30(9–10):1707–1727.

Gandhi, A., Navarro, S., and Rivers, D. (2011). On the identification of production functions: How heterogeneous is productivity? Technical Report 669206180, University of Western Ontario.

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.

Gopinath, G., Kalemli-Özcan, , Karabarbounis, L., and Villegas-Sanchez, C. (2017). Capital Allocation and Productivity in South Europe. *The Quarterly Journal of Economics*, 132(4):1915–1967.

Griliches, Z. and Mairesse, J. (1998). Production functions: The search for identification. In Strøm, S., editor, *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, pages 2413–2415, 2417, 2421. Cambridge University Press, Cambridge, U.K.

Guerron-Quintana, P., Inoue, A., and Kilian, L. (2013). Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 4(2):197–229.

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.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics*, 57(2):189–202.

Karabarbounis, L. and Neiman, B. (2018). Accounting for factorless income. *NBER Macroeconomics Annual*, 33:167–228.

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. (2009a). Inference on subsets of parameters in GMM without assuming identification. Working paper (updated version), Brown University.

Kleibergen, F. and Mavroeidis, S. (2009b). Weak instrument robust tests in gmm and the new keynesian phillips curve. *Journal of Business & Economic Statistics*, 27(3):293–311.

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.

Loecker, J. D. and Scott, P. T. (2017). Estimating market power: Evidence from the us brewing industry. *NBER Working Paper*, (24423).

Marschak, J. and Andrews, W. H. (1944). Random simultaneous equations and the theory of production. *Econometrica*, 12(3):143–205.

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.

Mavroeidis, S. (2010). Monetary policy rules and macroeconomic stability: Some new evidence. *American Economic Review*, 100(1):491–503.

Melitz, M. J. and Redding, S. J. (2014). Heterogeneous firms and trade. *Handbook of International Economics*, 4:1–54.

Mertens, M. (2019). Labour market power and the distorting effects of international trade. Technical Report 2/2019, Leibniz-Institut für Wirtschaftsforschung Halle (IWH), Halle (Saale).

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

Nason, J. M. and Smith, G. W. (2008). Identifying the new keynesian phillips curve. *Journal of Applied Econometrics*, 23(5):525–551.

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.

Rothenberg, T. J. (1971). Identification in parametric models. *Econometrica*, 39(3):577–591.

Ruge-Murcia, F. J. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control*, 31(8):2599–2636.

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.

Stock, J. H. and Yogo, M. (2005). Testing for weak instruments in linear iv regression. In Andrews, D. W. K. and Stock, J. H., editors, *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pages 80–108. Cambridge University Press, Cambridge, UK.

Syverson, C. (2019). Macroeconomics and market power: Facts, potential explanations, and open questions. *Brookings Papers on Economic Activity*, 2019(Spring):317–371.

Traina, J. (2018). Is aggregate market power increasing? production trends using financial statements. *Stigler Center Working Paper*, (17).

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. (2009). On estimating firm-level production variables to control for unobservables. *Economics Letters*, 104:112–114.

Yeh, C., Macaluso, C., and Hershbein, B. (2022). Monopsony in the us labor market. *American Economic Review*, 112(7):2099–2138.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics*, 86(3):797–810.

Tables

Table 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 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 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 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 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 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.

Figures

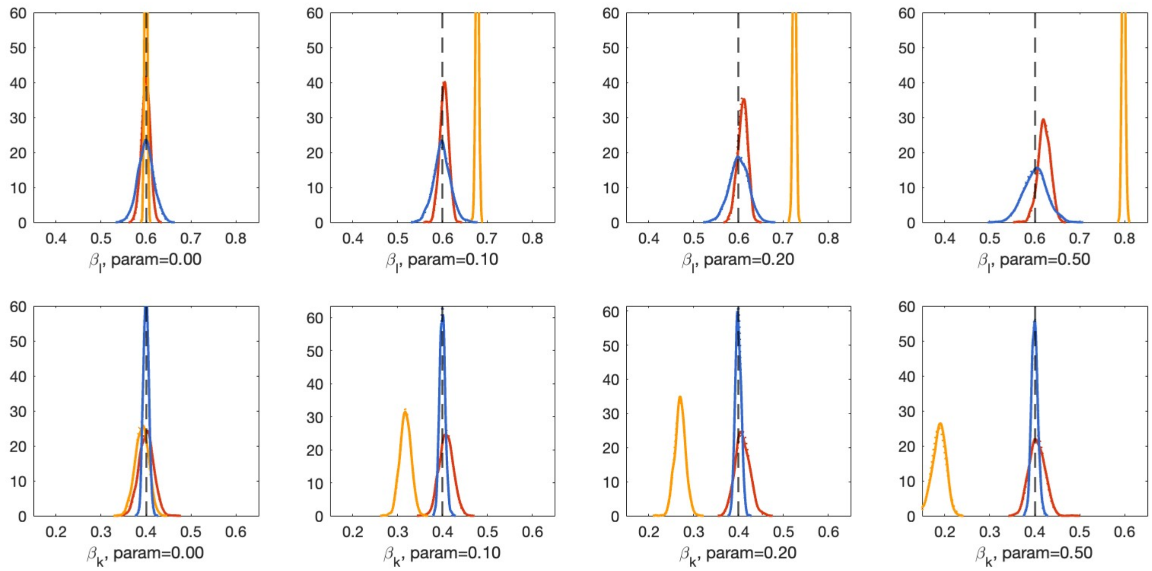

Figure 1: 1000 replications. $\sigma_{\xi_l} = 0.37$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value. Empirical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

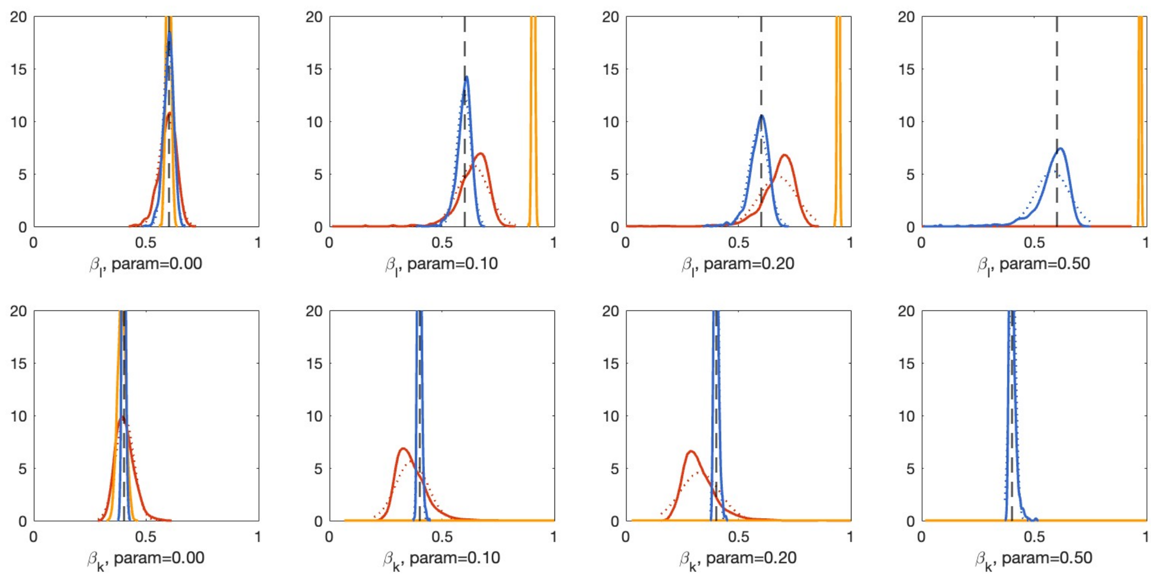

Figure 2: 1000 replications. $\sigma_{\xi_l} = 0.10$. Empirical densities of the estimates for β_l (top row) and β_k (bottom row) across measurement–error levels 0%, 10%, 20%, 50% (left to right). Methods: ACF (red), LP (orange), CUE (blue). The dashed vertical line marks the true parameter value. Empirical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

Appendix A Proofs

A.1 Proof of Proposition 1

We verify Assumptions 4–5 of Chen et al. (2014) and then invoke their Theorem 7. Denote as $\|\cdot\|$ the Euclidean distance.

Assumption 4 (Fréchet derivative linear and bounded). By the expressions for m'_β and m'_Γ given earlier, the differential $m'(\theta-\theta_0) = m'_\beta(\beta-\beta_0) + m'_\Gamma(\Gamma-\Gamma_0)$ is linear. It is bounded on the natural spaces $B := L^2(w_{it-1})$ and $N'_\Gamma := L^2(\widetilde{\omega}_{it-1})$ because, by the L^2 contraction of conditional expectation and (i)–(ii),

$$\|m'_\Gamma l\|_B \le \|l(\widetilde{\omega}_{it-1})\|_{L^2}, \qquad \|m'_\beta v\|_B \le \Big(\big(\mathbb{E}\|x_{it}\|^2\big)^{1/2} + \|\Gamma'_0\|_\infty \big(\mathbb{E}\|x_{it-1}\|^2\big)^{1/2}\Big)\|v\|.$$

(Rank condition). Assumption (iii) states exactly that Π is nonsingular.

Assumption 5 (uniform smoothness of m'_β). Fix $\varepsilon > 0$. Take neighborhoods $\mathcal{B}_r = \{\beta : \|\beta-\beta_0\| \le r\}$ and $\mathcal{N}_\Gamma^\delta = \{\Gamma : \sup_v |\Gamma'(v) - \Gamma'_0(v)| \le \delta\}$. We need that for some r, δ

$$\sup_{\Gamma\in\mathcal{N}_\Gamma^\delta} \left(\mathbb{E}\left[\sup_{\beta\in\mathcal{B}_r} \|m'_\beta(\beta,\Gamma) - m'_\beta(\beta_0,\Gamma_0)\|^2\right]\right)^{1/2} < \varepsilon$$

Define $V(\beta,\Gamma) := \Gamma'(\widetilde{\omega}_{it-1} + x_{it-1}^\top(\beta_0-\beta))x_{it-1}$. By the triangle inequality we can write

$$\begin{aligned}\|m'\beta(\beta,\Gamma) - m'\beta(\beta_0,\Gamma_0)\|^2 &= \|\mathbb{E}[V(\beta,\Gamma) \mid w_{it-1}] - \mathbb{E}[V(\beta_0,\Gamma_0) \mid w_{it-1}]\|^2 \\ &\le 2\|\mathbb{E}[V(\beta,\Gamma) - V(\beta,\Gamma_0)|\ w_{it-1}]\|^2 \\ &\quad + 2\|\mathbb{E}[V(\beta,\Gamma_0) - V(\beta_0,\Gamma_0)]|\ w_{it-1}]\|^2.\end{aligned}$$

We can bound the first term using Jensen's inequality. For any β

$$\mathbb{E}[\|\mathbb{E}[V(\beta,\Gamma) - V(\beta,\Gamma_0)|\ w_{it-1}]\|^2] \le \mathbb{E}[(\Gamma'(v_\beta) - \Gamma'_0(v_\beta))^2\|x_{it-1}\|^2]$$

for $v_\beta = \widetilde{\omega}_{it-1} + x_{it-1}^\top(\beta_0-\beta)$. Hence, by the above inequality we have

$$\mathbb{E}\left[\sup_{\beta\in\mathcal{B}_r}\|\mathbb{E}[V(\beta,\Gamma) - V(\beta,\Gamma_0)|\ w_{it-1}]\|^2\right] \le \mathbb{E}\left[\sup_{\beta\in\mathcal{B}_r}(\Gamma'(v_\beta) - \Gamma'_0(v_\beta))^2\|x_{it-1}\|^2\right]$$

By finite moments in ii) and definition of $\mathcal{N}_\Gamma^\delta$ we have the following term is bounded and can be chosen arbitrarily small by fixing the right δ

$$\sup_{\Gamma \in \mathcal{N}_\Gamma^\delta} \left(\mathbb{E} \left[\sup_{\beta \in \mathcal{B}_r} \|\mathbb{E}[V(\beta, \Gamma) - V(\beta, \Gamma_0)| \; w_{it-1}]\|^2 \right] \right) \leq \delta^2 \mathbb{E}\|x_{it-1}\|^2$$

For the second term we have by Jensen for any β

$$\mathbb{E}[\|\mathbb{E}[V(\beta, \Gamma_0) - V(\beta_0, \Gamma_0)|w_{it-1}]\|^2] \leq \mathbb{E}\big(|\Gamma_0'(v_\beta) - \Gamma_0'(\widetilde{\omega}_{it-1})|^2 \, \|x_{it-1}\|^2\big) \, ,$$

By the Lipschitz property in (i), with Lipschitz constant C,

$$\mathbb{E}[\|\mathbb{E}[V(\beta, \Gamma_0) - V(\beta_0, \Gamma_0)|w_{it-1}]\|^2] \leq C^2 \, \mathbb{E}[| \, x_{it-1}^\top (\beta_0 - \beta)|^2 \|x_{it-1}\|^2]$$

By Cauchy-Schwarz $|x^\top \Delta\beta|^2 \;\leq\; \|x\|^2 \, \|\Delta\beta\|^2$, hence, by the above inequality

$$\mathbb{E} \left[\sup_{\beta \in \mathcal{B}_r} \|\mathbb{E} \left[V(\beta, \Gamma_0) - V(\beta_0, \Gamma_0)|w_{it-1} \right] \|^2 \right] \leq C^2 \, \mathbb{E} \left[\sup_{\beta \in \mathcal{B}_r} \|\beta_0 - \beta\|^2 \|x_{it-1}\|^4 \right]$$

By finite moments in ii) and definition of $\mathcal{B}_r$ we have the following term is bounded and can be chosen arbitrarily small by fixing the right r

$$\sup_{\Gamma \in \mathcal{N}_\Gamma^\delta} \left(\mathbb{E} \left[\sup_{\beta \in \mathcal{B}_r} \|\mathbb{E}[V(\beta, \Gamma_0) - V(\beta_0, \Gamma_0)| \; w_{it-1}]\|^2 \right] \right) \leq C^2 r^2 \, \mathbb{E} \left[\|x_{it-1}\|^4 \right]$$

Combining the two terms, and taking square root we can write

$$\sup_{\Gamma \in \mathcal{N}_\Gamma^\delta} \left(\mathbb{E}\Big[\sup_{\beta \in \mathcal{B}_r} \|m'\beta(\beta, \Gamma) - m'\beta(\beta_0, \Gamma_0)\|^2 \Big] \right)^{1/2} \;\leq\; \big(2 \, \delta^2 \, \mathbb{E}\|x\|^2 + 2 \, C^2 \, r^2 \, \mathbb{E}\|x\|^4\big)^{1/2}.$$

Since we can choose r, δ arbitrarily, we have uniform smoothness, as required. Thus Assumption 5 holds.

With Assumptions 4–5 verified and (iii), Theorem 7 of Chen et al. (2014) yields local identification of β_0.

A.2 Proof of Theorem 2

Note that by condition ii) the Jacobian is dependent on n. For ease of notation we denote $\mathcal{J}_n(\theta_0) = \mathcal{J}_n$. Since f_n is twice continuously differentiable,

$$\sqrt{n} f_n(\theta) = \sqrt{n} f_n(\theta_0) + \sqrt{n} \, \partial f_n(\theta_0)/\partial\theta^\top \, (\theta - \theta_0) + \frac{1}{2}\sqrt{n} \, (\theta - \theta_0)^\top \left(\partial^2 f_n(\theta_0)/\partial\theta\partial\theta^\top \right) (\theta - \theta_0)$$

Asymptotically, the second term can be written as

$$\sqrt{n}\,\partial f_n(\theta_0)/\partial\theta^\top\,(\theta-\theta_0) \xrightarrow{d} \sqrt{n}\mathcal{J}_\beta(\beta-\beta_0)+\sqrt{n}\mathcal{J}_\rho(\rho-\rho_0),$$

and the last term as

$$\frac{1}{2}\sqrt{n}\,(\theta-\theta_0)^\top\left(\partial^2 f_n(\theta_0)/\partial\theta\partial\theta^\top\right)(\theta-\theta_0) \xrightarrow{d} \frac{1}{2}\sqrt{n}(\rho-\rho_0)^2\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} z_{it-1}x_{it-1}^\top\right]$$

Plugging in for ρ a root-n consistent estimator $\hat{\rho}_n$, such that $\sqrt{n}(\hat{\rho}_n-\rho_0) \xrightarrow{d} \psi_\rho$, gives the following limiting representation

$$\sqrt{n}f_n(\beta,\hat{\rho}_n) \xrightarrow{d} \psi_f + C(\beta-\beta_0) + r\psi_\rho$$

which results by applying condition (i) for the first term, plugging in for $\mathcal{J}_\beta$ and $\mathcal{J}_\rho$ in the second term using conditions (ii) and realizing that the last term is $o_p(1)$, since $\hat{\rho}_n$ is root-n consistent, and thus $\sqrt{n}(\hat{\rho}_n-\rho_0)^2 = O_p(n^{-1/2})$.

Appendix 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. 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 $h = \tau + 1$,

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

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} used in the Monte Carlo.

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

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 4. We use moment vector

$$f_{it}(\theta) = z_{it}\left(y_{it} - \alpha - \beta_k k_{it} - \beta_l l_{it} - \rho(y_{it-1} - \alpha - \beta_k k_{it-1} - \beta_l l_{it-1})\right), \tag{26}$$

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