Identification of Production Function Parameters with Noisy Proxy Variables: A Semiparametric Control-Function Analysis

Jorge de la Cal Medina*

November 26, 2025

1 Introduction

Proxy-variable methods such as Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015) address the simultaneity problem in production function estimation by assuming that investment or intermediate input choices are strictly increasing in firm productivity. This "scalar unobservable" condition implies that productivity can be perfectly recovered from observables, enabling control-function identification strategies.

Empirically, however, proxy variables are noisy. Input choices reflect not only productivity but also frictions, adjustment costs, shocks, and measurement error. This paper shows that full recoverability of productivity is not necessary for identification of the production-function coefficients. We allow intermediate-input demand to include an additive noise term, so that productivity is only partially predictable given the proxy. Our framework permits a realistic form of misspecification in the proxy structure while still enabling identification and estimation of structural parameters.

The structural coefficients β in the production function are identified via a semiparametric conditional moment restriction. The key idea is to exploit the fact that the proxy remains informative about productivity, even if it does not perfectly reveal it. Our approach uses the framework of Chen et al. (2014), treating the evolution of productivity as a nuisance function and applying orthogonalization to partial out its influence. Local identification is achieved

*University of Amsterdam. Email: j.delacalmedina@uva.nl

whenever an orthogonalized Jacobian matrix has full column rank, isolating variation in inputs not captured by the proxy index.

Recent work by Doraszelski and Li (2025) also relaxes the scalar–unobservable condition. They work within the canonical OP/LP/ACF proxy–variable framework and study what happens when the invertibility of the proxy fails, for example because of imperfect competition or unobserved demand shocks. They show how to modify the moments to eliminate the bias induced by non–invertibility.

Our paper is similar in that it allows productivity to be only partially predictable from the proxy and uses orthogonal moments with flexible first–step estimation, but differs by developing a semiparametric identification result that treats the productivity forecast function as a nonparametric nuisance.

We develop a semiparametric identification and estimation strategy based on orthogonal moments that are locally robust to first-stage estimation error in the spirit of Chen et al. (2014). Identification does not rely on invertibility, completeness, or full support. Instead, it requires mean independence and a forecast-sufficiency condition on the proxy, making the framework well suited to empirical settings with measurement error or frictions in input adjustment.

We propose an estimation method that follows the Neyman orthogonalization framework of Chernozhukov et al. (2022), constructing moment functions that are orthogonal to the nuisance tangent space associated with nonparametrically estimated regressions. This guarantees that small estimation errors in the first stage do not affect the large-sample distribution of the second-stage estimator.

2 Model

We briefly introduce the control function approach following OP/LP/ACF. Consider the following production function in logs for firm i at time t:[1]

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

y is the log of output, x is the vector of the log inputs, which includes a constant to account for an intercept and transformations of labor and capital, ε denotes unexpected productivity shocks, and ω captures firm-specific productivity. Note that implicitly we assume a

[1] This specification accommodates the commonly used Cobb-Douglas and translog production functions

"value-added specification", intermediate inputs m do not enter (1) directly but serve as a proxy for productivity, as we will see below.[2] Latin letters denote observed variables; Greek letters denote variables unobserved to the researcher. We are interested in identifying and estimating the production function parameter β based on a panel of firms $i = 1, \ldots, N$ over time periods $t = 1, \ldots, T$; with $N \to \infty$ and fixed T.

Behind the identification strategy is a discrete time model of optimizing firms. At time t, firm i chooses inputs conditional on observing a productivity draw ω_{it}. At the end of the period output y_{it} is produced depending on the chosen inputs and to a noise term ε_{it}, unexpected productivity. We allow for part of the input vector x_{it} to be predetermined at $t-1$, denoted at x^p_{it} and part to be flexibly determined at t, denoted as x^f_{it}, so $x_{it} = (x^p_{it}, x^f_{it})$. The firm also flexibly chooses upon intermediate inputs m_{it}. Hence the choice at t is upon next period's predetermined inputs, x^p_{it+1}, and this period's flexible inputs and intermediate inputs, x^f_{it}, m_{it}. OP/LP/ACF typically treat capital as predetermined and labor as flexible.

2.1 Model Assumptions

The model is assumed to satisfy the following assumptions. The identifying moment conditions will be based on the previously explained timing assumptions. It is thus helpful to introduce the information $\mathcal{I}_{it}$, which includes all variables known to firm i at t.

Assumption 1. *$\mathcal{I}_{it}$ includes all current and past realizations of ω_{it}, $\{\omega_{i\tau}\}_{\tau \le t} \subset \mathcal{I}_{it}$, and moreover satisfies $\mathbb{E}[\varepsilon_{it}|\mathcal{I}_{it}] = 0$.*

Consequently, every input decision the firm chooses as a function of ω_{it}, is also contained in $\mathcal{I}_{it}$. Thus $(x^p_{it+1}, x^f_{it+1}, m_{it}) \in \mathcal{I}_{it}$.

Assumption 2. *The productivity process ω_{it} follows a first-order Markov process:*

$$g(\omega_{it+1} \mid \mathcal{I}_{it}) = g(\omega_{it+1} \mid \omega_{it}), \tag{2}$$

where $g(\cdot \mid \omega_{it})$ denotes the conditional density, which is known to firms and is stochastically increasing in ω_{it}.

Assumptions 1–2 imply that ω_{it} can be written as its conditional expectation at time $t-1$

[2] For a discussion of the relationship between gross output and value-added production functions, see **?**, **?**, and **?**.

plus an innovation term, with a slight abuse of notation using $g(\cdot)$ for the conditional mean:

$$\omega_{it} = g(\omega_{it-1}) + \xi_{it}, \tag{3}$$

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

The fundamental assumption of the control function approach is that firm productivity ω_{it} can be proxied by observables. Specifically, it is assumed that there exists a demand function for a flexibly chosen intermediate input, m_{it}, which depends on productivity and predetermined inputs:

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

Inverting $h(\omega, x^p)$ with respect to ω yields a control function in terms of observables,

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

The canonical OP/LP/ACF framework imposes the so-called *scalar unobservable assumption*, under which the firm's choice of m_{it} is fully determined by ω_{it}, conditional on the predetermined input x^p_{it}. Equivalently, productivity is perfectly recoverable from the proxy variables:

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

This assumption implies that the proxy variables (m_{it}, x^p_{it}) are perfectly correlated with the productivity term and contain no independent variation. 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 such proxy errors, see Hu et al. (2020).

To relax this overly restrictive assumption, we allow for an unexplained component in the intermediate-input demand function. This admits measurement error in m and unobserved heterogeneity in firms' intermediate input decisions. We will show that we don't require exact recovery of productivity from observables for identification; instead, identification relies on *mean independence* of the resulting forecast errors with respect to the firm's information set, as formalized in Assumption 3 below. This point is also made in Doraszelski and Li (2025).

Assumption 3. *Intermediate input demand function with error*

$$m_{it} = h(\omega_{it}, x_{it}^p, \nu_{it}), \tag{4}$$

$h(\omega, x^p, \nu)$ is strictly monotone in ω.

The error ν_{it} may capture, for example, measurement error, adjustment frictions, or taste shocks in intermediate-input choice. Since the mapping $\omega \mapsto h(\omega, x_{it}^p, \nu_{it})$ is invertible, we can write

$$\omega_{it} = h^{-1}(m_{it}, x_{it}^p, \nu_{it}). \tag{5}$$

However, since ν_{it} is unobserved to the econometrician, productivity cannot be recovered from (m_{it}, x_{it}^p) alone, and the scalar-unobservable condition of OP/LP/ACF fails. Setting $\nu_{it} = 0$ collapses Assumption 3 to the usual scalar-unobservable case in which ω_{it} is perfectly recoverable from (m_{it}, x_{it}^p). Note that setting $\nu_{it} = 0$ collapses Assumption 3 to the usual scalar-unobservable case of OP/LP/ACF.

Let w_{it} be the vector of inputs chosen at by firm i at t, $w_{it} = ({x_{it+1}^p}', {x_{it}^f}', {m_{it}}')'$. Moreover define the observables, by the econometrician, at time t as

$$z_{it} = (w_{it}', w_{it-1}', \ldots, w_{i1}')'$$

The moment conditions implied by Assumptions 1–3 are defined with respect to the firm's information set $\mathcal{I}_{it}$. To link them to the researcher's observables, we impose the following assumption. Note that it holds immediately if z_{it} is $\mathcal{I}_{it}$-measurable, that is, if $z_{it} \subset \mathcal{I}_{it}$.

Assumption 4 (Researcher's information set)**.** *For all i, t and for $s \leq t$,*

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

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 $\Psi(z) = \mathbb{E}[\omega_{it} \mid z_{it} = z]$, with which we can define the index $\widetilde{\omega}_{it} := \Psi(z_{it})$. We then have

$$\mathbb{E}[\omega_{it} \mid z_{it}] \;=\; \widetilde{\omega}_{it}.$$

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

For identification, 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 $u \mapsto \mathbb{E}[\omega_{it+1} \mid \widetilde{\omega}_{it} = u]$; while g remains the structural law of motion in (3).

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.[3]

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 linear $g(\omega)$, as in the previous section satisfies Assumption 5 immediately.

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

$$\mathbb{E}[\omega_{it+1} \mid z_{it}] = \Gamma(\tilde{\omega}_{it}),$$

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

[3]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.

3 Identification

3.1 Identification

The control–function strategy proceeds in two steps.

First stage

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

$$y_{it} = x_{it}^\top \beta + \Psi(z_{it}) + \eta_{it} + \varepsilon_{it}. \tag{6}$$

Assumption 4 implies

$$\mathbb{E}[\eta_{it} + \varepsilon_{it} \mid z_{it}] = 0. \tag{7}$$

Because x_{it} may also enter $\Psi(\cdot)$ linearly, β is not identified from (6) 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}).[4]

Second Stage

By Assumption 5, we can write (1) as

$$y_{it} = x_{it}^\top \beta + \Gamma(\widetilde{\omega}_{it-1}) + \zeta_{it} + \varepsilon_{it}. \tag{8}$$

Assumptions 4–5 imply

$$\mathbb{E}[\zeta_{it} + \varepsilon_{it} \mid z_{it-1}] = 0. \tag{9}$$

[4] See Chernozhukov et al. (2022) for a general treatment of nonparametric identification in conditional moment models.

This implies the following second-stage regression function

$$\begin{aligned}\mathbb{E}[y_{it} \mid z_{it-1}] &= \mathbb{E}[x_{it}^\top \beta \mid z_{it-1}] \; + \; \Gamma\big(\widetilde{\omega}_{it-1}\big) \\ &= \mathbb{E}[x_{it}^\top \beta \mid z_{it-1}] \; + \; \Gamma\big(\Phi(z_{it-1}) - x_{it-1}^\top \beta\big),\end{aligned} \tag{10}$$

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 (8)-(9) 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 $m'_\Gamma(\Gamma_0 - \Gamma)$

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

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

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

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

Define 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). *Suppose Assumptions 1–5 hold. Assume further that* Γ_0 *is continuously differentiable with Lipschitz derivative, and that the regressors satisfy* $\mathbb{E}\|x_{it}\|^4 < \infty$ *for all* t. *If the orthogonalized Jacobian matrix* Π *is nonsingular, then the structural parameter* β_0 *is locally identified.*

Sketch of proof. We apply Theorem 7 of Chen et al. (2014), which establishes local identification under three main conditions: (i) linear and bounded Fréchet differentiability of the moment function, (ii) full rank of the orthogonalized Jacobian Π, and (iii) uniform smoothness of the score operator with respect to the structural parameter. The rank condition is directly assumed to hold, so we verify the other two conditions (see Appendix for full proof).

Fréchet differentiability. The moment function $m(\beta, \Gamma)$ is Fréchet differentiable at (β_0, Γ_0) with derivative

$$m'(\theta - \theta_0) = m'_\beta(\beta - \beta_0) + m'_\Gamma(\Gamma - \Gamma_0),$$

which is linear by construction. Boundedness follows from the Lipschitz continuity of Γ'_0 and the moment condition $\mathbb{E}\|x_{it}\|^4 < \infty$.

Uniform smoothness. Uniform smoothness of $m'_\beta(\beta, \Gamma)$ follows from the Lipschitz continuity of Γ'_0 and the fourth-moment condition on x_{it}, together with standard envelope and domination arguments over small neighborhoods of (β_0, Γ_0).

These conditions imply the local identification of β_0 by Theorem 7 of Chen et al. (2014). □

4 Estimation

This section presents two estimators for the structural parameter β: a two-step semiparametric GMM estimator and a one-step orthogonalized estimator following Chernozhukov et al. (2022). Both approaches are based on the conditional moment restriction developed in Section 3, but they differ in how they treat the infinite-dimensional nuisance functions.

In both cases we use GMM (Hansen, 1982). For a given parameter vector θ and corresponding sample moment function

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

where $r_{it}(\theta)$ is the residuals and vector z_{it} collects the instruments, the GMM estimator is defined as

$$\hat{\theta} := \arg\min_{\theta \in \Theta} \left\| f_{nT}(\theta) \right\|_{W}^{2},$$

with $\|v\|_W^2 := v^\top W v$ denoting the quadratic form induced by a symmetric positive semi-definite weighting matrix W.

4.1 Two-Step Estimation via Series GMM

The first approach estimates the conditional expectation Φ_0 and the productivity evolution function Γ in separate steps, followed by GMM estimation of the structural parameters.

Step 1. We estimate the conditional expectation function

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

nonparametrically, for example by series regression. Let $\hat{\Phi}_0$ denote the resulting estimator.

Step 2. Using (8), define the second-step residual

$$r_{2,it}(\beta, \Gamma) := y_{it} - x_{it}^\top \beta - \Gamma\Big(\hat{\Phi}_0(z_{it-1}) - x_{it-1}^\top \beta\Big).$$

We approximate Γ by a sieve

$$\Gamma_J(v; \gamma) := \sum_{j=1}^{J} \gamma_j \phi_j(v),$$

where $\{\phi_j\}_{j=1}^J$ is a given set of basis functions, γ is a J-dimensional coefficient vector, and $J = J_{nT}$ is the sieve dimension. Substituting the sieve approximation yields

$$r_{2,it}(\beta, \gamma) := y_{it} - x_{it}^\top \beta - \Gamma_J\Big(\hat{\Phi}_0(z_{it-1}) - x_{it-1}^\top \beta; \gamma\Big)\,.$$

Define the second-step sample moment

$$f_{2,nT}(\beta, \gamma) := \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=2}^{T} r_{2,it}(\beta, \gamma)\, z_{it-1}.$$

Collecting the parameters as $\theta_2 := (\beta^\top, \gamma^\top)^\top$, the two-step series GMM estimator is

$$\hat{\theta}_2 := (\hat{\beta}_2, \hat{\gamma}) := \arg\min_{\theta_2} \big\| f_{2,nT}(\theta_2) \big\|_W^2.$$

Under standard regularity conditions on the sieve approximation, moment bounds, and dependence across (i, t), $\hat{\beta}_2$ is consistent and asymptotically normal.[5]

4.2 One-Step Estimation via Orthogonal Moments

The second approach avoids separate estimation of Φ_0 and Γ_0 by directly targeting the composite function

$$\Lambda_0(z_{it-1}; \beta) := \Gamma_0\big(\Phi_0(z_{it-1}) - x_{it-1}^\top \beta\big)\,,$$

which captures the expected productivity component conditional on observables and indexed by the structural parameter β. At the true parameter, $\Lambda_0(z_{it-1}; \beta_0) = \mathbb{E}[\omega_{it} \mid z_{it-1}]$.

We treat $\Lambda_0(\cdot; \beta)$ as a nuisance function and approximate it nonparametrically as a function of (z_{it-1}, β). Let $\hat{\Lambda}(z_{it-1}; \beta)$ denote a nonparametric estimator of $\Lambda_0(z_{it-1}; \beta)$. This yields the one-step residual

$$r_{1,it}(\beta, \hat{\Lambda}) := y_{it} - x_{it}^\top \beta - \hat{\Lambda}(z_{it-1}; \beta).$$

Consider $\hat{\Lambda}(z_{it-1}; \beta)$ to be the nonparametric regression of $y_{it} - x_{it}^\top \beta$ on z_{it-1}, for example, a series or kernel estimator. Then, $\hat{\Lambda}(z_{it-1}; \beta)$ is an explicit function of the residuals $y_{it} - x_{it}^\top \beta$ and the regressors z_{it-1} conditional on β, so it can be updated jointly with the GMM criterion rather than in a separate preliminary step. [6] We can thus omit the dependence on $\hat{\Lambda}$ and

[5] See, e.g., Newey and McFadden (1994) for series GMM asymptotics.

[6] In practice, the choice of tuning parameters (e.g. bandwidths, sieve dimension, or other regularization

write the one-step residual as $r_{1,it}(\beta)$.

The one-step sample moment is given by

$$f_{1,nT}(\beta) := \frac{1}{n(T-1)} \sum_{i=1}^{n} \sum_{t=2}^{T} r_{1,it}(\beta)\, z_{it-1}$$

The one-step estimator of β is then defined as

$$\hat{\beta}_1 := \arg\min_{\beta} \left\| f_{1,nT}(\beta) \right\|_W^2.$$

Under standard regularity conditions and under the Neyman-orthogonality conditions of Chernozhukov et al. (2022), the orthogonalized GMM estimator $\hat{\beta}_1$ is consistent and asymptotically normal.[7]

Neyman orthogonality. Neyman orthogonality makes the GMM moments locally insensitive to errors in the nonparametric (or high-dimensional) nuisance estimates: first-stage estimation error only enters at second order, so $\hat{\beta}_1$ is $\sqrt{n}$-consistent and asymptotically normal even when the nuisance is estimated at a slower rate.

We verify that the one-step moments are first-order Neyman orthogonal with respect to the nuisance function $\Lambda_0(\cdot;\beta)$. Denote the population moment of the one-step estimator as

$$m_1(\beta, \Lambda) := \mathbb{E}\big[r_{1,it}(\beta, \Lambda)\, z_{it-1}\big],$$

with $m_1(\beta_0, \Lambda_0) = 0$ at the true parameter (β_0, Λ_0). Consider a local perturbation of Λ_0 in the direction h, $\Lambda_0 + \tau h$, then

$$m_1(\beta, \Lambda_0 + rh) = \mathbb{E}\big[\, \big(y_{it} - x_{it}^{\top}\beta - \Lambda_0(z_{it-1};\beta) - \tau\, h(z_{it-1})\big)\ z_{it-1}\big].$$

Define the nuisance tangent space

$$\mathcal{H} := \big\{ h \in L^2(z_{it-1}) : \mathbb{E}[h(z_{it-1})\, z_{it-1}] = 0 \big\}.$$

parameters) must be made in advance, possibly by data–driven rules such as cross–validation. Under the usual rate conditions ensuring consistency of $\hat{\Lambda}$, and given Neyman orthogonality of the moment condition, this tuning affects only higher–order terms and does not change the first–order asymptotic distribution of $\hat{\beta}_1$; see Chernozhukov et al. (2022).

[7] See, e.g., Chernozhukov et al. (2022) for asymptotics of locally robust semiparametric GMM estimators with orthogonal moments.

The Gateaux derivative in direction $h \in \mathcal{H}$ is

$$\frac{\partial}{\partial \tau} m_1(\beta, \Lambda_0 + \tau h)\bigg|_{\tau=0} = -\,\mathbb{E}\big[h(z_{it-1})\, z_{it-1}\big].$$

For any $h \in \mathcal{H}$ we have $\mathbb{E}[h(z_{it-1})\, z_{it-1}] = 0$, thus the moment $m_1(\beta, \Lambda)$ is first-order Neyman orthogonal with respect to the nuisance function $\Lambda_0(\cdot; \beta)$ in the sense of Bonhomme et al. (2024).

5 Conclusion

This paper provides a semiparametric identification analysis for production function models estimated using proxy variables. Exact invertibility of the proxy is unnecessary. Instead, identification relies on conditional mean-independence, forecast sufficiency, and variation in inputs orthogonal to the index determining the nuisance function. The orthogonalized Jacobian matrix fully characterizes the identifying content of the control-function moments.

References

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

Bonhomme, S., Jochmans, K., and Weidner, M. (2024). A neyman-orthogonalization approach to the incidental parameter problem. Working paper. arXiv:2412.10304 [econ.EM].

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

Chernozhukov, V., Escanciano, J. C., Ichimura, H., Newey, W. K., and Robins, J. (2022). Locally robust semiparametric estimation. *Econometrica*, 90(4):1501–1535.

Doraszelski, U. and Li, L. (2025). Identification of production functions with non-invertible proxies. Working paper, 2025.

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

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.

Levinsohn, J. and Petrin, A. (2003). Estimating production functions using intermediate inputs to control for unobservables. *Review of Economic Studies*, 70(2):317–341.

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

Appendix

A.1 Proof of Proposition 1

We verify the sufficient conditions for local identification in Theorem 7 of Chen et al. (2014) by establishing their Assumptions 4–5. Let $\|\cdot\|$ denote the Euclidean norm.

Fréchet differentiability (Assumption 4). The Gateaux derivative of the moment condition $m(\beta, \Gamma)$ is given by

$$m'(\theta - \theta_0) = m'_\beta(\beta - \beta_0) + m'_\Gamma(\Gamma - \Gamma_0),$$

where

$$m'_\beta v = -\mathbb{E}[x_{it} - \Gamma_0'(\widetilde{\omega}_{it-1})x_{it-1} \mid z_{it-1}], \quad m'_\Gamma l = -\mathbb{E}[l(\widetilde{\omega}_{it-1}) \mid z_{it-1}].$$

Linearity is immediate. Boundedness follows from the L^2 contraction of conditional expectation and the assumptions:

$$\|m'_\Gamma l\| \leq \|l(\widetilde{\omega}_{it-1})\|_{L^2}, \quad \|m'_\beta v\| \leq \left(\sqrt{\mathbb{E}\|x_{it}\|^2} + \|\Gamma_0'\|_\infty \sqrt{\mathbb{E}\|x_{it-1}\|^2}\right) \|v\|.$$

Rank condition. By assumption, the orthogonalized Jacobian matrix Π is nonsingular.

Uniform smoothness (Assumption 5). Fix $\varepsilon > 0$ and define neighborhoods

$$\mathcal{B}_r = \{\beta : \|\beta - \beta_0\| \leq r\}, \quad \mathcal{N}_\Gamma^\delta = \{\Gamma : \sup_v |\Gamma'(v) - \Gamma_0'(v)| \leq \delta\}.$$

Define

$$V(\beta, \Gamma) := \Gamma'(\widetilde{\omega}_{it-1} + x_{it-1}^\top(\beta_0 - \beta))x_{it-1}.$$

Then

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

For the first term, Jensen's inequality and the definition of $\mathcal{N}_\Gamma^\delta$ yield:

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

For the second term, using Lipschitz continuity of Γ_0' (constant C) and Cauchy–Schwarz:

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

Combining both terms,

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

Choosing δ, r small enough ensures this is below ε.

Conclusion. With Assumptions 4–5 verified and Π nonsingular, Theorem 7 of Chen et al. (2014) implies that β_0 is locally identified.

Remark. The condition $\mathbb{E}[\|x_{it-1}\|^4] < \infty$ is required for bounding the second term in the uniform smoothness argument. Thus, it is a necessary assumption.