Production Function Estimation without Invertibility: Imperfectly Competitive Environments and Demand Shocks*

Ulrich Doraszelski

University of Pennsylvania†

Lixiong Li

Johns Hopkins University‡

March 20, 2025

Abstract

We revisit the proxy variable approach to production function estimation. We show that the invertibility assumption at its heart is testable. We characterize what goes wrong if invertibility fails and what can still be done. We show that rethinking how the estimation procedure is implemented either eliminates or mitigates the bias that arises if invertibility fails. Furthermore, we show how a modification of the procedure ensures Neyman orthogonality and enhances efficiency and robustness.

1 Introduction

Production functions are one of the oldest concepts in economics (von Thünen 1842, Wicksteed 1894, Cobb & Douglas 1928). As a description of the relationship between inputs and output,

*We thank Dan Ackerberg and Steve Berry for helpful discussions and comments and Alexander Whitefield, Vivian Zhang, and Wanxi Zhou for excellent research assistance.

†Wharton School, Email: doraszelski@wharton.upenn.edu

‡Department of Economics, Email: lixiong.li@jhu.edu

they are of interest in and of themselves and also a vehicle for measuring productivity and technical change (Solow 1957). More recently, production functions have become a key input into estimating markups and markdowns in order to assess firms' market power (Hall 1988, De Loecker & Warzynski 2012). This production approach to markup estimation recognizes that, given an estimate of the production function, the markup can be recovered from the firm's cost minimization problem. To estimate the production function a large and rapidly growing literature relies on the procedure initially developed by Olley & Pakes (1996) and further advanced by Levinsohn & Petrin (2003) and Ackerberg, Caves & Frazer (2015) (henceforth, OP, LP, and ACF).

The OP/LP/ACF procedure starts from the observation due to Marschak & Andrews (1944) that there is an endogeneity problem in estimating production functions. This problem arises because the decisions that the firm makes regarding its inputs depend on its productivity, which is unobserved by the econometrician. To resolve this problem, OP turn it on its head and use the firm's decisions that are observed by the econometrician to infer its productivity. The resulting proxy variable approach to production function estimation has quickly become standard practice.

The inversion from observables to productivity at the heart of the OP/LP/ACF procedure requires that firms with the same productivity make the same choices. Yet, there is no reason to believe this is the case, e.g., if these firms face different demands in the output market. The current paper therefore revisits the OP/LP/ACF procedure if invertibility fails. It characterizes what goes wrong in the OP/LP/ACF procedure and what can still be done.

The observation that the OP/LP/ACF procedure cannot accommodate unobserved demand heterogeneity has been made before. Foster, Haltiwanger & Syverson (2008) put it as follows:

> ... idiosyncratic demand shocks make the proxies functions of both technology and demand shocks, thereby inducing a possible omitted variable bias. Put simply, proxy methods require a one-to-one mapping between plant-level productivity and the observables used to proxy for productivity. This mapping breaks down if other unobservable plant-level factors besides productivity drive changes in the observable proxy. (p. 403)

Indeed, to establish invertibility, OP rule out that firms face different demands and abstract from competition between firms.[1] LP assume a perfectly competitive industry where firms act as price takers and thus face the same horizontal demand curve (p. 322 and Appendix A). Based on their results, ACF start from scalar unobservable and strict monotonicity assumptions, and almost all subsequent literature simply imposes invertibility as a high-level assumption. For this reason, it may not have fully appreciated the demanding nature of the invertibility assumption.

[1] OP assume that any profitability differences across firms are due to differences in their capital stocks and productivities (p. 1273), thereby ruling out that firms face different demands. Limiting the state variables in the firm's investment policy to its own capital stock and productivity moreover abstracts from competition between firms (see also Lemma 3 and Theorem 1 in Pakes 1994).

Unobserved demand heterogeneity is likely to be empirically important. The large literatures on demand estimation and productivity analysis highlight the considerable heterogeneity in demand that remains even after controlling for detailed product attributes (Berry, Levinsohn & Pakes 1995) or honing in on (nearly) homogenous products (Foster et al. 2008). The problem is compounded by the fact that, in imperfectly competitive environments, a firm's decisions in equilibrium depend on its own productivity as well as on the productivities of its rivals. Hence, imperfectly competitive environments require jointly inverting the decisions of all firms for the productivities of all firms. This many-to-many inverse may not exist (Biondi 2022) or it may be too high-dimensional to be practical (Ackerberg & De Loecker 2024). Moreover, a firm's rivals are partially or completely unobserved in typical datasets used for production function estimation. The decisions of its rivals have thus to be thought of as shocks to the demand the focal firm faces.

Invertibility can fail for reasons other than unobserved demand heterogeneity. Changes in firm conduct due to mergers and acquisitions or switches from competition to collusion are best thought of as changes to the ownership matrix. Invertibility can fail if these changes in firm conduct are partially or completely unobserved by the econometrician. Finally, invertibility can fail if there is unobserved variation across firms or time in input prices, investment opportunities, or financial constraints.

Taken together, the invertibility assumption is especially challenging for the production approach to markup estimation for two reasons. First, the demand a firm faces in the output market is a fundamental determinant of the markup that its charges. Second, estimating markups is arguably most important in imperfectly competitive environments and after changes in firm conduct.

This paper makes five contributions. First, we propose tests for invertibility. Our tests exploit that if invertibility fails, then productivity becomes a hidden state in a Markov model. The literature on Kalman and particle filters in dynamic systems shows that the best guess for the hidden state uses the entire history of the observables as opposed to just their current value. Invertibility can therefore be tested by including lags of the observables in the regression in the first step of the OP/LP/ACF procedure. Implementing our tests on an unbalanced panel of Spanish manufacturing firms and a balanced panel of US manufacturing industries, we strongly reject invertibility.

Building on Doraszelski & Jaumandreu (2024), we therefore characterize the consequences of a failure of invertibility for the OP/LP/ACF procedure. If invertibility fails, then the prediction of quantity from the first step contains an error. Because this prediction is used to control for lagged productivity in the GMM estimation in the second step, the lagged prediction error enters into the conditional moment. This invalidates capital as an instrument, resulting in biased estimates. However, because the prediction error is by construction mean independent of the observables

used in the first step, other instruments such as lagged capital, labor, and materials remain valid.

Turning from what goes wrong if invertibility fails to what can still be done, our second contribution is to provide a necessary and sufficient condition for the moment condition in the second step of the OP/LP/ACF procedure to hold for the true production function and some (not necessarily the true) law of motion for productivity. This condition calls for rethinking how the OP/LP/ACF procedure is implemented. In particular, it compels us to ensure that any instrument used in the second step is appropriately included in the regression in the first step. Due to the lag structure of the model, this means including the lead of capital in addition to its current value in the regression. We provide a series of examples where this simple expedient suffices to satisfy our necessary and sufficient condition.

Beyond these examples, our necessary and sufficient condition may be violated. While this results in biased estimates, our third contribution is to show that rethinking how the OP/LP/ACF procedure is implemented mitigates the bias. Specifically, we show that the moment condition in the second step implicitly incorporates a first-order bias correction provided any instrument used in the second step is appropriately included in the regression in the first step.

Our fourth contribution is to explicitly include the bias correction in the second step of the OP/LP/ACF procedure. We show that the resulting modified moment condition enjoys a property known as Neyman orthogonality (Neyman 1959). While our modification is as easy to implement as the original OP/LP/ACF procedure, the benefit of Neyman orthogonality is considerable. Under some conditions, it notably renders the asymptotic distribution of the GMM estimator in the second step invariant to the quality of the prediction of productivity from the first step. This is especially beneficial if the regression in the first step uses a rich set of covariates. In finite samples, Neyman orthogonality improves the performance of the GMM estimator. A Monte Carlo exercise shows that this improvement can be substantial.

Our fifth and final contribution is to provide a diagnostic to assess the sensitivity of the estimates to the size of the prediction error in the first step. While our diagnostic has some similarities to the sensitivity measure in Andrews, Gentzkow & Shapiro (2017), it is not local to the true model and directly informative about the estimated model. Our diagnostic is neither necessary nor sufficient for no bias. However, a small value of the diagnostic provides assurance that small changes in the prediction error do not dramatically alter the estimates.

In sum, this paper revisits the OP/LP/ACF procedure if the invertibility assumption fails. Invertibility fails if there are demand shocks or in imperfectly competitive environments with partially or completely unobserved rivals or changes in firm conduct. Whether invertibility fails can be tested. A failure of invertibility can have a substantial, detrimental impact on the estimates.

Fortunately, much can still be done. We provide a necessary and sufficient condition for the moment condition in the second step of the OP/LP/ACF procedure to hold for the true production

function. This condition compels us to ensure that any instrument used in the second step is appropriately included in the regression in the first step. We show that this simple change either eliminates or mitigates the bias. Going a step further, we modify the moment condition in the second step to endow it with Neyman orthogonality. Finally, we provide a diagnostic to assess the sensitivity of the estimates to the size of the prediction error that arises in the first step of the OP/LP/ACF procedure if invertibility fails.

The remainder of the paper is organized as follows. In Section 2, we recall the setup and the OP/LP/ACF procedure. In Section 3, we develop tests for invertibility. In Section 4, we recall that a failure of invertibility invalidates capital as an instrument, resulting in biased estimates. We develop a necessary and sufficient condition for the moment condition in the second step to hold for the true production function. We show that ensuring that any instrument used in the second step is appropriately included in the regression in the first step either eliminates or mitigates the bias. In Section 5, we modify the moment condition in the second step to endow it with Neyman orthogonality. In Section 6, we conduct a Monte Carlo exercise to illustrate what goes wrong in the OP/LP/ACF procedure if invertibility fails and what can still be done. In Section 7, we provide a diagnostic to assess the sensitivity of the estimates to the size of the prediction error. We conclude in Section 8.

2 Setup and OP/LP/ACF procedure

Firm i in period t produces output Q_{it} with inputs K_{it} and V_{it} according to the production function

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

where lower case letters denote logs. Capital k_{it} is a predetermined input that is chosen in period $t-1$ whereas v_{it} is freely variable and decided on in period t after the firm observes its productivity ω_{it}.[2] Productivity follows a first-order Markov process with law of motion

$$\omega_{it} = \mathrm{E}\left[\omega_{it}|\omega_{it-1}\right] + \xi_{it} = g(\omega_{it-1}) + \xi_{it}, \tag{2}$$

where the productivity innovation ξ_{it} is by construction mean independent of lagged productivity ω_{it-1} and further assumed to be mean independent of any variable included in the firm's information set in period $t-1$. The disturbance ε_{it} sits between the firm's output q_{it} as recorded in the data and the output $q^*_{it} = q_{it} - \varepsilon_{it} = f(k_{it}, v_{it}) + \omega_{it}$ that the firm planned on when it decided on the variable input v_{it}. It can be interpreted as measurement error or as an unanticipated

[2] While k_{it} and v_{it} may be vectors, we think of them as scalars for simplicity. The variable input v_{it} may accordingly be interpreted as a composite of labor and materials such as cost of goods sold.

shock to output (OP, pp. 1273–1274) and is assumed to be mean independent of the inputs and other included variables as formalized below.[3] While the econometrician observes actual output q_{it} and the inputs k_{it} and v_{it}, productivity ω_{it}, the disturbance ε_{it}, and planned output q^*_{it} remain unobserved.

Invertibility. The literature following OP relies on invertibility. Invertibility assumes that there exists a function $\omega_{it} = h(x_{it})$ that maps observables $x_{it} = (k_{it}, v_{it}, \ldots)$ into productivity ω_{it}. This is equivalently to

$$\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}.$$

The observables included in x_{it} depend on the decision of the firm that is being inverted.[4] We remain agnostic about which variables are included in x_{it}, aside from imposing, without loss of generality, that the inputs k_{it} and v_{it} are included.

OP/LP/ACF procedure. Estimation proceeds in two steps. Step 1 assumes $\mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$ and flexibly or nonparametrically estimates the conditional expectation

$$\mathrm{E}\left[q_{it}|x_{it}\right] = \mathrm{E}\left[f(k_{it}, v_{it}) + \omega_{it} + \varepsilon_{it}|\, x_{it}\right] = f(k_{it}, v_{it}) + \mathrm{E}\left[\omega_{it}|x_{it}\right] = f(k_{it}, v_{it}) + \omega_{it} = q^*_{it}, \qquad (3)$$

where the first equality uses equation (1), the second-to-last equality uses invertibility, and the last equality uses the definition of planned output q^*_{it}. Because $\mathrm{E}\left[q_{it}|x_{it}\right] = q^*_{it}$, step 1 separates actual output q_{it} into planned output q^*_{it} and the disturbance $\varepsilon_{it} = q_{it} - q^*_{it}$.

Step 2 of the OP/LP/ACF procedure assumes $\mathrm{E}\left[\xi_{it} + \varepsilon_{it}|\, z_{it}\right] = 0$ for instruments $z_{it} = (k_{it}, k_{it-1}, v_{it-1}, \ldots)$ and estimates the parameters $\theta = (\theta_f, \theta_g)$ in the production function f and the law of motion g by GMM.[5] Capital k_{it} is a valid instrument because it is a predetermined input, and the lagged inputs k_{it-1} and v_{it-1} are valid instruments because the productivity innovation ξ_{it} is mean independent of any variable included in the firm's information set in period $t-1$. We remain agnostic regarding additional instruments in z_{it}.

Using equations (1) and (2) and the definition of planned output q^*_{it}, the assumption $\mathrm{E}\left[\xi_{it} + \varepsilon_{it}|\, z_{it}\right] = 0$ implies

$$\mathrm{E}[q_{it} - f(k_{it}, v_{it}) - g(\omega_{it-1})|z_{it}] = \mathrm{E}[q_{it} - f(k_{it}, v_{it}) - g(q^*_{it-1} - f(k_{it-1}, v_{it-1}))|z_{it}] = 0. \qquad (4)$$

Substituting $\mathbb{E}[q_{it-1}|x_{it-1}]$ from step 1 for q^*_{it-1}, step 2 proceeds by estimating the parameters θ

[3] Mundlak & Hoch (1965) refer to ω_{it} and ε_{it} as the transmitted, respectively, untransmitted component of productivity. The untransmitted component may include machine breakdowns, labor actions, supply chain disruptions, and power outages that are not anticipated by the firm. The interpretation as measurement error accommodates serial correlation in the disturbance ε_{it}.

[4] OP invert the firm's demand for investment whereas LP and ACF invert its demand for materials.

[5] For notational convenience we suppress θ in much of what follows.

from the moment condition

$$\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f(k_{it-1}, v_{it-1})\right)|z_{it}\right] = 0. \tag{5}$$

Throughout the remainder of the paper, we follow OP, LP, and ACF and maintain $\mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$ for observables $x_{it} = (k_{it}, v_{it}, \ldots)$ and $\mathrm{E}\left[\xi_{it} + \varepsilon_{it}|\, z_{it}\right] = 0$ for instruments $z_{it} = (k_{it}, k_{it-1}, v_{it-1}, \ldots)$. To avoid burdensome notation, we ask the reader to understand equality involving random variables and conditional expectations as equality almost surely.

3 Tests for invertibility

If invertibility fails and $\mathrm{E}\left[\omega_{it}|x_{it}\right] \neq \omega_{it}$, then productivity ω_{it} becomes a hidden state in a Markov model. The literature on Kalman and particle filters in dynamic systems shows that the best guess for the unobservables uses the entire history of the observables as opposed to just their current value. Our tests for invertibility are based on this intuition.

Our first proposition shows that invertibility implies a testable mean-independence restriction:

Proposition 1. *If* $\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}$ *and* $\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right]$, *then* $\mathrm{E}\left[q_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[q_{it}|x_{it}\right]$.

While step 1 of the OP/LP/ACF procedure assumes $\mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$ for observables $x_{it} = (k_{it}, v_{it}, \ldots)$ and step 2 assumes $\mathrm{E}\left[\xi_{it} + \varepsilon_{it}|z_{it}\right] = 0$ for instruments $z_{it} = (k_{it}, k_{it-1}, v_{it-1}, \ldots)$, this does not quite imply $\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right]$. However, the interpretation of the disturbance ε_{it} as an unanticipated shock to output that is outside the firm's information set in period t implies $\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$.[6] The proof of Proposition 1 is straightforward:

Proof. If $\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}$, then

$$q_{it} = f(k_{it}, v_{it}) + \omega_{it} + \varepsilon_{it} = f(k_{it}, v_{it}) + \mathrm{E}\left[\omega_{it}|x_{it}\right] + \varepsilon_{it}.$$

$\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right]$ therefore implies $\mathrm{E}\left[q_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[q_{it}|x_{it}\right]$. □

Our second proposition shows that invertibility implies a testable conditional-independence restriction:

Proposition 2. *If* $\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}$ *and* $\varepsilon_{it} \perp\!\!\!\perp x_{it-1}|\, x_{it}$, *then* $q_{it} \perp\!\!\!\perp x_{it-1}|x_{it}$.

[6] Under the interpretation of the disturbance ε_{it} as measurement error, the assumption $\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right]$ holds if x_{it-1} does not have predictive power for ε_{it} beyond x_{it}.

Going beyond the first-order implication of $\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}$ in Proposition 1 calls for a stronger assumption on the disturbance ε_{it}. The assumption $\varepsilon_{it} \perp\!\!\!\perp x_{it-1} | \, x_{it}$ in Proposition 2 means that any possible dependence between ε_{it} and x_{it-1} is through x_{it}. It is implied by $\varepsilon_{it} \perp\!\!\!\perp (x_{it}, x_{it-1})$ and implies $\mathrm{E}\left[\varepsilon_{it}|x_{it}, x_{it-1}\right] = \mathrm{E}\left[\varepsilon_{it}|x_{it}\right]$. The proof of Proposition 2 parallels the proof of Proposition 1 and is therefore omitted.

We implement our tests for invertibility on an unbalanced panel of Spanish manufacturing firms from 1990 to 2006 (Encuesta Sobre Estrategias Empresariales) and a balanced panel of US manufacturing industries from 1958 to 2018 (NBER-CES). In both cases, we reject invertibility by a wide margin. We provide further details on the data and the tests in Appendix A. In the next section, we turn to the consequences of a failure of invertibility for production function estimation.

4 Failure of invertibility and validity of OP/LP/ACF moment condition

If invertibility fails and $\mathbb{E}\left[\omega_{it}|x_{it}\right] \neq \omega_{it}$, then a prediction error $\zeta_{it} = \omega_{it} - \mathbb{E}\left[\omega_{it}|x_{it}\right]$ arises that is generally non-zero. By construction, the prediction error satisfies $\mathbb{E}\left[\zeta_{it}|x_{it}\right] = 0$.

To see how the presence of the prediction error ζ_{it} affects the OP/LP/ACF procedure, recall that the conditional expectation estimated in step 1 is

$$\mathrm{E}\left[q_{it}|x_{it}\right] = f(k_{it}, v_{it}) + \mathrm{E}\left[\omega_{it}|x_{it}\right] + \mathrm{E}[\varepsilon_{it}|x_{it}] = f(k_{it}, v_{it}) + \omega_{it} - \zeta_{it} = q^*_{it} - \zeta_{it}, \tag{6}$$

where the second-to-last equality uses $\mathrm{E}[\varepsilon_{it}|x_{it}] = 0$ and the definition of the prediction error ζ_{it}, and the last equality uses the definition of planned output q^*_{it}. Equation (6) provides an alternative interpretation of the prediction error as the difference between planned output and its prediction in step 1: $\zeta_{it} = q^*_{it} - \mathrm{E}[q_{it}|x_{it}]$.

Substituting $\mathrm{E}[q_{it-1}|x_{it-1}]$ from equation (6), the left-hand side of moment condition (5) in step 2 becomes

$$\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)|z_{it}\right]. \tag{7}$$

Because of the lagged prediction error ζ_{it-1}, conditional moment (7) is generally non-zero. To see this more clearly, consider the special case of an $AR(1)$ process for productivity. If $g(\omega_{it-1}) = \rho\omega_{it-1}$, then conditional moment (7) evaluates to

$$\mathrm{E}\left[\xi_{it} + \varepsilon_{it} + \rho\zeta_{it-1}|z_{it}\right] = \rho\mathrm{E}\left[\zeta_{it-1}|z_{it}\right]. \tag{8}$$

Recall that $x_{it-1} = (k_{it-1}, v_{it-1}, \ldots)$ and $z_{it} = (k_{it}, k_{it-1}, v_{it-1}, \ldots)$. The lagged inputs k_{it-1} and v_{it-1} hence remain valid instruments because $\mathrm{E}\left[\zeta_{it-1}|x_{it-1}\right] = 0$. However, capital k_{it} may no longer

be a valid instrument as $\mathrm{E}\left[\zeta_{it-1}|k_{it}\right] \neq 0$, as previously pointed out by Doraszelski & Jaumandreu (2024). This is because the firm chooses k_{it} in period $t-1$ with knowledge of ω_{it-1}. What remains of lagged productivity after controlling for the lagged observables x_{it-1} may therefore be correlated with k_{it}.

The above discussion foreshadows part of the solution to the problem: adding the lead of capital k_{it+1} to the observables x_{it} in step 1 of the OP/LP/ACF procedure ensures that $\mathrm{E}\left[\zeta_{it-1}|k_{it}\right] = 0$ and hence the validity of capital k_{it} as an instrument in step 2. This simple expedient amounts to rethinking how the OP/LP/ACF procedure is implemented. The remaining question is to what extent it suffices beyond the special case of an $AR(1)$ process for productivity.

The following theorem provides a complete answer to this question in the form of a necessary and sufficient condition.

Theorem 1. *Moment condition (5) in step 2 of the OP/LP/ACF procedure holds for the true production function f^0 and some law of motion $\tilde{g}$ if and only if*

$$\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] = \mathrm{E}\left[\tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)|\, z_{it}\right], \tag{9}$$

where g^0 is the true law of motion.

We prove Theorem 1 before parsing it:

Proof. Recall that the true production function f^0 and the true law of motion g^0 satisfy moment condition (4).
"Only if" part: Suppose moment condition (5) holds for the true production f^0 and some law of motion $\tilde{g}$. Then we have

$$\begin{aligned} 0 &= \mathrm{E}\left[q_{it} - f^0(k_{it}, v_{it}) - \tilde{g}\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f^0(k_{it-1}, v_{it-1})\right)\big|\, z_{it}\right] \\ &= \mathrm{E}\left[q_{it} - f^0(k_{it}, v_{it}) - \tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)\big|\, z_{it}\right], \end{aligned}$$

where the last equality uses equation (6). Subtracting from equation (4) implies condition (9).
"If" part: Suppose condition (9) holds. Substituting into equation (4), we have

$$\begin{aligned} 0 &= \mathrm{E}\left[q_{it} - f^0(k_{it}, v_{it}) - \tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)\big|\, z_{it}\right] \\ &= \mathrm{E}\left[q_{it} - f^0(k_{it}, v_{it}) - \tilde{g}\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f^0(k_{it-1}, v_{it-1})\right)\big|\, z_{it}\right], \end{aligned}$$

where the last equality uses equation (6). Hence, moment condition (5) holds for f^0 and $\tilde{g}$. □

Theorem 1 covers the invertibility assumption at the heart of the literature following OP as a special case. If $\mathrm{E}\left[\omega_{it}|x_{it}\right] = \omega_{it}$, then condition (9) becomes $\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] = \mathrm{E}\left[\tilde{g}\left(\omega_{it-1}\right)|z_{it}\right]$

and is therefore satisfied with $\tilde{g} = g^0$. Importantly, however, condition (9) can be satisfied even if invertibility fails and $\mathrm{E}\left[\omega_{it}|x_{it}\right] \neq \omega_{it}$.

To understand condition (9), note that it becomes $\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] = \tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)$ if $x_{it-1} \subsetneq z_{it}$. This is difficult to satisfy as the left-hand side is a function of z_{it} whereas the right-hand side is a function of $x_{it-1} \subsetneq z_{it}$.

We therefore set $x_{it-1} \supseteq z_{it}$ in what follows. This amounts to rethinking how the OP/LP/ACF procedure is implemented, as foreshadowed by the special case of an $AR(1)$ process for productivity. In particular, because k_{it} is included in z_{it} in step 2, we must include k_{it+1} in x_{it} in step 1.

Setting $x_{it-1} \supseteq z_{it}$, we further examine condition (9) through a series of examples:

Example 1 (linear law of motion, $AR(1)$ process)**.** If $g^0(\omega_{it-1}) = \rho\omega_{it-1}$, then $\mathrm{E}\left[\rho\omega_{it-1}|z_{it}\right] = \mathrm{E}\left[\rho\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]|z_{it}\right]$. Condition (9) is therefore satisfied with $\tilde{g}(\mathrm{E}[\omega_{it-1}|x_{it-1}]) = g^0(\mathrm{E}[\omega_{it-1}|x_{it-1}])$. ■

Remarkably, moment condition (5) in step 2 of the OP/LP/ACF procedure holds in Example 1 irrespective of the quality of the estimate of $\mathrm{E}\left[q_{it}|x_{it}\right]$ in step 1.

Example 2 (quadratic law of motion)**.** If $g^0(\omega_{it-1}) = \rho_1\omega_{it-1} + \rho_2\omega_{it-1}^2$ and $\mathrm{Var}\left(\zeta_{it-1}|z_{it}\right) = \sigma^2$, then

$$\mathrm{E}\left[\rho_1\omega_{it-1} + \rho_2\omega_{it-1}^2\middle| z\right] = \mathrm{E}\left[\rho_1\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right] + \rho_2\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]^2 + \rho_2\zeta_{it-1}^2\middle| z_{it}\right].$$

Condition (9) is therefore satisfied with $\tilde{g}(\mathrm{E}[\omega_{it-1}|x_{it-1}]) = g^0(\mathrm{E}[\omega_{it-1}|x_{it-1}]) + \rho_2\sigma^2$.[7] ■

Example 2 is less restrictive on the law of motion but more restrictive on the lagged prediction error than Example 1. The next example pushes this tradeoff further:

Example 3 (analytic law of motion)**.** If the Taylor series of g^0 around $\mathrm{E}[\omega_{it-1}|x_{it-1}]$ converges absolutely so that $g^0(\omega_{it-1}) = \sum_{j=0}^{\infty} \frac{g^{0,(j)}(\mathrm{E}[\omega_{it-1}|x_{it-1}])}{j!}\zeta_{it-1}^j$, where $g^{0,(j)}$ is the jth derivative of g^0, then

$$\begin{aligned}\mathrm{E}\left[\sum_{j=0}^{\infty}\frac{g^{0,(j)}(\mathrm{E}[\omega_{it-1}|x_{it-1}])}{j!}\zeta_{it-1}^j\middle| z_{it}\right] &= \mathrm{E}\left[\mathrm{E}\left[\sum_{j=0}^{\infty}\frac{g^{0,(j)}(\mathrm{E}[\omega_{it-1}|x_{it-1}])}{j!}\zeta_{it-1}^j\middle| x_{it-1}\right]\middle| z_{it}\right] \\ &= \mathrm{E}\left[\sum_{j=0}^{\infty}\frac{g^{0,(j)}(\mathrm{E}[\omega_{it-1}|x_{it-1}])}{j!}\mathrm{E}[\zeta_{it-1}^j|x_{it-1}]\middle| z_{it}\right].\end{aligned}$$

If $\mathrm{E}[\zeta_{it-1}^j|x_{it-1}] = \alpha_j$ for some constant α_j for all j, then condition (9) is satisfied in this example with $\tilde{g}(\mathrm{E}[\omega_{it-1}|x_{it-1}]) = \sum_{j=0}^{\infty}\frac{g^{0,(j)}(\mathrm{E}[\omega_{it-1}|x_{it-1}])}{j!}\alpha_j$.[8] ■

[7] We can relax $\mathrm{Var}\left(\zeta_{it-1}|z_{it}\right) = \sigma^2$ to $\mathrm{Var}\left(\zeta_{it-1}|z_{it}\right) = h(\mathrm{E}[\omega_{it-1}|x_{it-1}])$ for some function h. In this case, condition (9) holds with $\tilde{g}(\mathrm{E}[\omega_{it-1}|x_{it-1}]) = g^0(\mathrm{E}[\omega_{it-1}|x_{it-1}]) + \rho_2 h(\mathrm{E}[\omega_{it-1}|x_{it-1}])$.

[8] Condition (9) remains valid if we relax the condition that $\mathrm{E}[\zeta_{it-1}^j|x_{it-1}] = \alpha_j$ for some constant α_j for all j to $\zeta_{it-1} \perp\!\!\!\perp z_{it}|\,\mathrm{E}[\omega_{it-1}|x_{it-1}]$.

Beyond these examples, condition (9) may be violated even if $x_{it-1} \supseteq z_{it}$.[9] While this results in biased estimates, setting $x_{it-1} \supseteq z_{it}$ mitigates the bias. If $x_{it-1} \supseteq z_{it}$, then $\mathrm{E}\left[\omega_{it-1}|z_{it}\right] = \mathrm{E}\left[\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]|z_{it}\right]$ by the law of iterated expectations and hence $\mathrm{E}\left[\zeta_{it-1}|z_{it}\right] = \mathrm{E}\left[\omega_{it-1} - \mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]|\, z_{it}\right] = 0$. This mean independence of the lagged prediction error ζ_{it-1} from the instruments z_{it} intuitively eliminates a first-order source of bias. The following theorem formalizes this intuition:

Theorem 2. *If $x_{it-1} \supseteq z_{it}$ and g is differentiable, then moment condition (5) in step 2 of the OP/LP/ACF procedure implicitly incorporates a first-order bias correction:*

$$\begin{aligned}\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f(k_{it-1}, v_{it-1})\right)|z_{it}\right] = \mathrm{E}\Big[q_{it} - f(k_{it}, v_{it}) \\ - g\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right) - g'\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1}\big|z_{it}\Big], \quad (10)\end{aligned}$$

where g' is the first derivative of g.

We note that equation (10) holds for any production function f and any law of motion g for which the conditional moments exist. The proof of Theorem 2 is in Appendix B.

To appreciate Theorem 2, recall that moment condition (4) remains valid even if invertibility fails. Because lagged planned output q^*_{it-1} is unobserved, moment condition (4) is infeasible for estimation. Step 2 of the OP/LP/ACF procedure therefore proceeds by substituting $\mathrm{E}\left[q_{it-1}|x_{it-1}\right]$ from step 1 for q^*_{it-1}. If invertibility fails, then this substitution introduces the lagged prediction error $\zeta_{it-1} = q^*_{it-1} - \mathrm{E}\left[q_{it-1}|x_{it-1}\right]$ into moment condition (5) (see conditional moment (7)). Theorem 2 shows that setting $x_{it-1} \supseteq z_{it}$ adjusts moment condition (5) toward the valid but infeasible moment condition (4).

Indeed, comparing the term $g\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f(k_{it-1}, v_{it-1})\right)$ in moment condition (5) to the term $g\left(q^*_{it-1} - f(k_{it-1}, v_{it-1})\right)$ in moment condition (4) yields

$$\begin{aligned} & g\left(\mathrm{E}\left[q_{it-1}|x_{it-1}\right] - f(k_{it-1}, v_{it-1})\right) - g\left(q^*_{it-1} - f(k_{it-1}, v_{it-1})\right) \\ = \; & g\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right) - g\left(q^*_{it-1} - f(k_{it-1}, v_{it-1})\right) \\ = \; & -g'\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1} + o(\zeta_{it-1}), \quad (11)\end{aligned}$$

where the last equality uses a first-order Taylor expansion of g around $q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})$ and $o(\zeta_{it-1})$ denotes the higher-order remainder. Equation (10) in Theorem 2 shows that if $x_{it-1} \supseteq z_{it}$, then moment condition (5) implicitly accounts for the term $-g'\left(q^*_{it-1} - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1}$. Moment condition (5) thus differs from the valid but infeasible moment condition (4) at most by higher-order terms. Put differently, setting $x_{it-1} \supseteq z_{it}$ provides a first-order bias correction in step

[9] Testing condition (9) is empirically challenging. A necessary condition for it to be satisfied is that moment condition (5) in step 2 of the OP/LP/ACF procedure holds at the estimated production function and law of motion. This can be tested with an overidentification test that corrects for the plug-in nature of the OP/LP/ACF procedure and clustering in the data.

2 of the OP/LP/ACF procedure.

Because setting $x_{it-1} \supseteq z_{it}$ provides a first-order bias correction in step 2 of the OP/LP/ACF procedure, a violation of condition (9) must be of second order relative to size of the prediction error. The following theorem formalizes this intuition:

Theorem 3. *Let $\mathcal{G}$ be a set of functions that includes the true law of motion g^0. If $x_{it-1} \supseteq z_{it}$ and g^0 is twice continuously differentiable, then*

$$\inf_{\tilde{g} \in \mathcal{G}} \left\| \mathrm{E}\left[g^0(\omega_{it-1}) | z_{it}\right] - \mathrm{E}\left[\tilde{g}\left(\mathrm{E}\left[\omega_{it-1} | x_{it-1}\right]\right) | z_{it}\right] \right\|_{L,1} \leq \tau \mathrm{Var}\left(\zeta_{it-1}\right),$$

where $\|\cdot\|_{L,1}$ is the L_1 norm, $\tau = \sup_{\omega_{it-1}} |g^{0\prime\prime}(\omega_{it-1})|$, and $g^{0\prime\prime}$ is the second derivative of g^0.

The proof of Theorem 3 is in Appendix B.

Theorem 3 ties the violation of condition (9) to the variance of the lagged prediction error ζ_{it-1}. Because $\mathrm{Var}(\zeta_{it-1}) = \mathrm{E}[(\omega_{it-1} - \mathrm{E}[\omega_{it-1}|x_{it-1}])^2]$, this variance gets smaller as the observables x_{it-1} get richer even beyond z_{it}. Theorem 3 therefore suggests to take a kitchen sink approach to the regression in step 1 of the OP/LP/ACF procedure and add as many relevant covariates as possible. Adding covariates that speak to demand conditions and firm conduct may be especially helpful.[10]

Adding covariates makes the assumption $\mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$ that underpins step 1 of the OP/LP/ACF procedure more demanding. We advocate replacing $x_{it} = (k_{it}, v_{it}, \ldots)$ in the original procedure by $x_{it} = (k_{it+1}, k_{it}, v_{it}, \ldots)$. In the original procedure, the assumption $\mathrm{E}\left[\varepsilon_{it}|x_{it}\right] = 0$ is justified by stipulating that ε_{it} is outside the firm's information set in period t. This justification extends to our approach as the firm is assumed to choose k_{it+1} in period t. However, it may be advisable to add current values or lags rather than leads of other covariates to x_{it} to guard against the possibility that the firm eventually learns ε_{it}.

Finally, we caution that adding covariates introduces a curse of dimensionality. The resulting larger mean squared error of the estimate of $\mathrm{E}[q_{it-1}|x_{it-1}]$ may counterbalance the benefit of the smaller $\mathrm{Var}(\zeta_{it-1})$ in finite samples. In the next section, propose a modification of the OP/LP/ACF procedure that addresses this problem.

Summary. Our analysis calls for rethinking how the OP/LP/ACF procedure is implemented. Theorem 1 makes a strong case for ensuring $x_{it-1} \supseteq z_{it}$. For a linear law of motion, the moment condition in step 2 of the OP/LP/ACF procedure holds for the true production function f^0 and

[10] The recent literature in fact proceeds along this line: "The materials demand function in our setting will take as arguments all state variables of the firm (...), including productivity, and all additional variables that affect a firm's demand for materials. These include firm location (...), output prices (...), product dummies (...), market shares (...), input prices (...), the export status of a firm (...), and the input (...) and output tariffs (...) that the firm faces on the product it produces" (De Loecker, Goldberg, Khandelwal & Pavcnik 2016, p. 466). Whether adding covariates restores invertibility can be assessed using the tests in Section 3.

the true law of motion g^0. For a nonlinear law of motion, recovering the true production function f^0 requires condition (9) to be satisfied. A violation of condition (9) results in biased estimates. In this case, Theorem 2 shows that setting $x_{it-1} \supseteq z_{it}$ provides a first-order bias correction.

Theorem 1 further suggests to be flexible in modeling the law of motion in step 2 as condition (9) is easier to satisfy if $\tilde{g}$ can come from a larger set $\mathcal{G}$. Theorem 3 finally suggests to take a kitchen sink approach to step 1.

Related literature. An alternative to relaxing the invertibility assumption is to forgo step 1 of the OP/LP/ACF procedure. Instead of substituting $\mathrm{E}\left[q_{it-1}|x_{it-1}\right]$ for q^*_{it-1} in moment condition (4), the dynamic panel approach to production function estimation pioneered by Blundell & Bond (2000) substitutes q_{it-1} and imposes the linear law of motion $g(\omega_{it-1}) = \rho\omega_{it-1}$ to obtain

$$\mathrm{E}[q_{it}-f(k_{it},v_{it})-g(q_{it-1}-f(k_{it-1},v_{it-1}))|z_{it}] = \mathrm{E}[q_{it}-f(k_{it},v_{it})-g(q^*_{it-1}-f(k_{it-1},v_{it-1}))-\rho\varepsilon_{it-1}|z_{it}].$$

Besides the assumption $\mathrm{E}\left[\xi_{it}+\varepsilon_{it}|z_{it}\right] = 0$ maintained in moment condition (4), the dynamic panel approach requires assuming $\mathrm{E}[\varepsilon_{it-1}|z_{it}] = 0$. In comparison, our Example 1 requires setting $x_{it-1} \supseteq z_{it}$ and assuming $\mathrm{E}[\varepsilon_{it-1}|x_{it-1}] = 0$ in step 1 of the OP/LP/ACF procedure. Either assumption can be justified by appealing to the firm's information set and the timing of decisions, as discussed above.

Hu, Huang & Sasaki (2020), Brand (2020), and Ponder (2021) generalize the dynamic panel approach from a linear to a nonlinear law of motion.[11] To illustrate the similarities and differences with our approach, consider the quadratic law of motion $g(\omega_{it-1}) = \rho_1\omega_{it-1}+\rho_2\omega^2_{it-1}$. Substituting q_{it-1} for q^*_{it-1} in moment condition (4) yields

$$\begin{aligned}\mathrm{E}[q_{it} - f(k_{it},v_{it}) - g(q_{it-1} - f(k_{it-1},v_{it-1}))|z_{it}] = \mathrm{E}[q_{it} - f(k_{it},v_{it}) \\ - g(q^*_{it-1} - f(k_{it-1},v_{it-1})) - \rho_1\varepsilon_{it-1} - 2\rho_2\varepsilon_{it-1}(q^*_{it-1} - f(k_{it-1},v_{it-1})) - \rho_2\varepsilon^2_{it-1}|z_{it}].\end{aligned}$$

The dynamic panel approach therefore requires assuming $\mathrm{E}[\varepsilon_{it-1}|z_{it}] = 0$, $\mathrm{E}[\varepsilon_{it-1}\omega_{it-1}|z_{it}] = 0$, and $\mathrm{Var}\left(\varepsilon_{it-1}|z_{it}\right) = \sigma^2$. In comparison, our Example 2 requires setting $x_{it-1} \supseteq z_{it}$ and assuming $\mathrm{E}[\varepsilon_{it-1}|x_{it-1}] = 0$ and $\mathrm{Var}(\zeta_{it-1}|z_{it}) = \sigma^2$. Both approaches are similar in that they impose exclusion restrictions on second-order moments of latent variables to ensure that the moment condition holds for the true production function.

This similarity extends to the polynomial law of motion $g(\omega_{it-1}) = \sum_{j=1}^{\infty}\rho_j\omega^j_{it-1}$. In this case, the dynamic panel approach requires assuming that $\mathrm{E}[\varepsilon^j_{it-1}|\omega_{it-1},z_{it}]$ is constant for all j.

[11] The estimation strategy in Hu et al. (2020) and Brand (2020) differs from the identification strategy. The latter builds on Hu & Schennach (2008) and relies on the existence of two conditionally independent proxies for productivity and the invertibility of an integral operator. The implied semiparametric maximum likelihood estimator involves significant computational challenges that preclude its implementation.

This restriction on the joint distribution of ε_{it-1} and ω_{it-1} may be unpalatable if ε_{it-1} and ω_{it-1} are interpreted as the untransmitted, respectively, transmitted component of productivity (see footnote 3).[12] In comparison, our Example 3 requires assuming that $\mathrm{E}[\zeta^j_{it-1}|x_{it-1}]$ is constant for all j.

The key difference is that our approach imposes restrictions on the lagged prediction error ζ_{it-1} rather than the lagged disturbance ε_{it-1}. Whereas the magnitude and properties of ε_{it-1} are inherent in the data generating process, the advantage of our approach is that the magnitude of ζ_{it-1} can be reduced by adding covariates to the regression in step 1 of the OP/LP/ACF procedure. Proceeding to add covariates may shrink $\mathrm{Var}(\zeta_{it-1}) = \mathrm{E}[(\omega_{it-1} - \mathrm{E}[\omega_{it-1}|x_{it-1}])^2]$ towards zero if we approach invertibility in the limit. This makes our approach particularly attractive for datasets with a rich set of observables.

5 Modification of OP/LP/ACF moment condition

To address the problem posed by a noisy estimate in step 1 of the OP/LP/ACF procedure, we propose explicitly incorporating the bias correction from Theorem 2. This modification ensures that estimation error in step 1 does not adversely affect the asymptotic distribution of the GMM estimator in step 2.

We think of $\mathrm{E}[q_{it-1}|x_{it-1}]$ as the true value of a nuisance parameter denoted by $e(x_{it-1})$. The finite-sample estimate of $e(x_{it-1})$ from step 1 of the OP/LP/ACF procedure generally deviates from the true value $\mathrm{E}[q_{it-1}|x_{it-1}]$. This deviation, in turn, adversely affects the asymptotic distribution of the GMM estimator in step 2 according to the sensitivity of the conditional moment to deviations of $e(x_{it-1})$ from $\mathrm{E}[q_{it-1}|x_{it-1}]$. To eliminate this adverse impact, we modify the moment condition in step 2 as follows:

$$\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1})\right) \right. \\ \left. - g'\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - e(x_{it-1})\right) \middle| z_{it}\right] = 0. \quad (12)$$

Consider first the case where $e(x_{it-1})$ equals its true value $\mathrm{E}[q_{it-1}|x_{it-1}]$. If $e(x_{it-1}) = \mathrm{E}[q_{it-1}|x_{it-1}]$ and $\mathrm{E}[\varepsilon_{it-1}|x_{it-1}, z_{it}] = 0$, then the left-hand side of moment condition (12) reduces to the right-hand side of equation (10) for any production function f and any law of motion g. Consequently, by Theorem 2, if $x_{it-1} \supseteq z_{it}$, then moment condition (12) coincides with moment condition (5) in step 2 of the OP/LP/ACF procedure. Moment condition (12) therefore has exactly the same identification properties as the original procedure in this case.

[12] If $x_{it-1} \subseteq z_{it}$ and invertibility holds, then $\mathrm{E}[\varepsilon^j_{it-1}|\omega_{it-1}, z_{it}] = \mathrm{E}[\varepsilon^j_{it-1}|z_{it}]$. While this facilitates interpreting the restriction, we focus on the case where invertibility fails.

However, if $e(x_{it-1})$ differs from its true value $\mathrm{E}[q_{it-1}|x_{it-1}]$, then moment condition (12) has the following property that moment condition (5) lacks:

Theorem 4 (Neyman orthogonality). *Define the set of functions* $\tilde{\mathcal{T}} = \{\delta(x_{it-1}) = e(x_{it-1}) - \mathrm{E}[q_{it-1}|x_{it-1}] : e \in \mathcal{T}\}$, *where* $\mathcal{T}$ *is the set of real-valued integrable functions of* x_{it-1}. *If* $x_{it-1} \supseteq z_{it}$ *and* g *has a bounded and continuous derivative, then the pathwise (or Gateaux) derivative of the left-hand side of moment condition* (12) *at* $e(x_{it-1}) = \mathrm{E}[q_{it-1}|x_{it-1}]$ *with respect to any direction* $\delta \in \tilde{\mathcal{T}}$ *is zero. That is, for any* $\delta \in \tilde{\mathcal{T}}$, *we have*

$$\frac{\partial}{\partial \lambda} \mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\right.$$
$$\left. - g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}] - \lambda\delta(x_{it-1})\right)\middle| z_{it}\right]\Big|_{\lambda=0} = 0.$$

The proof of Theorem 4 is in Appendix D.[13]

Theorem 4 shows that the left-hand side of moment condition (12) is not sensitive to small deviations of $e(x_{it-1})$ from its true value $\mathrm{E}[q_{it-1}|x_{it-1}]$. This property, known as Neyman orthogonality (Neyman 1959), has received renewed attention in the double-debiased machine learning literature. As discussed in Chernozhukov, Chetverikov, Demirer, Duflo, Hansen, Newey & Robins (2018) and Chernozhukov, Escanciano, Ichimura, Newey & Robins (2022), as long as the estimator for $e(x_{it-1})$ converges to $\mathrm{E}[q_{it-1}|x_{it-1}]$ at a rate faster than $N^{-1/4}$, Neyman orthogonality guarantees that the GMM estimator based on moment condition (12) is oracle efficient. This means that the asymptotic distribution of the GMM estimator is the same irrespective of whether we replace $e(x_{it-1})$ by its estimate from step 1 or by its true value $\mathrm{E}[q_{it-1}|x_{it-1}]$. As a byproduct, a practical advantage is that there is no longer any need to adjust standard errors to account for the first-step estimation following, e.g., Hahn, Liao & Ridder (2018).

In contrast to the general approaches to generating Neyman orthogonality using influence functions or score projections in Chernozhukov et al. (2022), moment condition (12) is Neyman orthogonal by construction. Because it avoids introducing high-dimensional nuisance parameters, our modification is as easy to implement as the original OP/LP/ACF procedure.

Theorem 4 allows for any integrable deviation and thus accommodates a wide range of estimation methods in step 1 of the OP/LP/ACF procedure. This includes traditional nonparametric methods such as sieve, kernel, and LASSO estimators, which produce smooth estimates, as well as modern machine learning techniques such as neural networks and random forests, which may produce non-smooth estimates. Intuitively, Neyman orthogonality means that small deviations of $e(x_{it-1})$ from its true value $\mathrm{E}[q_{it-1}|x_{it-1}]$ that can arise from finite-sample variance, functional approximation error in sieve estimation, local smoothing bias in kernel estimation, or regulariza-

[13] The assumption that g has a bounded derivative simplifies stating Theorem 4. As shown in Appendix D, it can be replaced by a much weaker assumption.

tion bias in LASSO and machine learning methods do not have a first-order impact on the GMM estimator based on moment condition (12).

Finally, we recommend ensuring $x_{it-1} \supseteq z_{it}$ when implementing our modification of the OP/LP/ACF procedure. This is because, without $x_{it-1} \supseteq z_{it}$, our modification provides a first-order bias correction on $e(x_{it-1})$ toward $\mathrm{E}[q_{it-1}|x_{it-1}, z_{it}]$ instead of $\mathrm{E}[q_{it-1}|x_{it-1}]$. To see this more clearly, we use the law of iterated expectations to rewrite moment condition (12) as

$$\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\right.$$
$$\left. - g'\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\left(\mathrm{E}[q_{it-1}|x_{it-1}, z_{it}] - e(x_{it-1})\right)\middle| z_{it}\right] = 0.$$

The conditional mean of the term in the second line is generally non-zero at $e(x_{it-1}) = \mathrm{E}[q_{it-1}|x_{it-1}]$ and the true production function f^0.

6 Monte Carlo exercise

Data generating process. We specify the CES production function

$$f(k_{it}, v_{it}) = \alpha_0 + \frac{\nu}{\rho} \ln\left(\alpha_k \exp(\rho k_{it}) + (1 - \alpha_k)\exp(\rho v_{it})\right),$$

where $\alpha_k \in (0, 1)$ is a distributional parameter, $\rho \leq 0$ and $\sigma = \frac{1}{1-\rho}$ is the elasticity of substitution, and $\nu > 0$ is returns to scale, and the disturbance $\varepsilon_{it} \sim N(0, \sigma_\varepsilon^2)$. We further specify the law of motion

$$g(\omega_{it-1}) = \mu_\omega + \rho_\omega \left((1 - \alpha_\omega)\omega_{it-1} + \frac{\alpha_\omega}{\zeta_\omega} \ln(\ln(1 + \exp(\zeta_\omega \omega_{it-1})))\right),$$

where $\rho_\omega \in (0, 1)$, $\alpha_\omega \in [0, 1]$, and $\zeta_\omega > 0$, and the productivity innovation $\xi_{it} \sim N(0, \sigma_\omega^2)$. Our particular interest lies in comparing the nonlinear process for $\alpha_\omega = 1$ with the Gaussian $AR(1)$ process for $\alpha_\omega = 0$. To facilitate this comparison we calibrate μ_ω and σ_ω^2 to hold fixed $\mathrm{E}[\omega_{it}]$ and $\mathrm{Var}(\omega_{it})$. Figure 1 illustrates the resulting distribution of productivity and overlays the law of motion in the two cases.

We specify the CES demand

$$q_{it}^* = \delta_{1it} - (1 + \exp(-\delta_{2it}))p_{it},$$

where p_{it} is the output price and $\delta = (\delta_{1it}, \delta_{2it})$ captures shocks to the demand the firm faces and unobserved rivals. This avoids having to specify the imperfectly competitive environment that the firm operates in.

We denote the price of capital as p_{it}^K and the price of the variable input as p_{it}^V. Assuming short-run profit maximization and equating marginal revenue with marginal cost determines v_{it} along

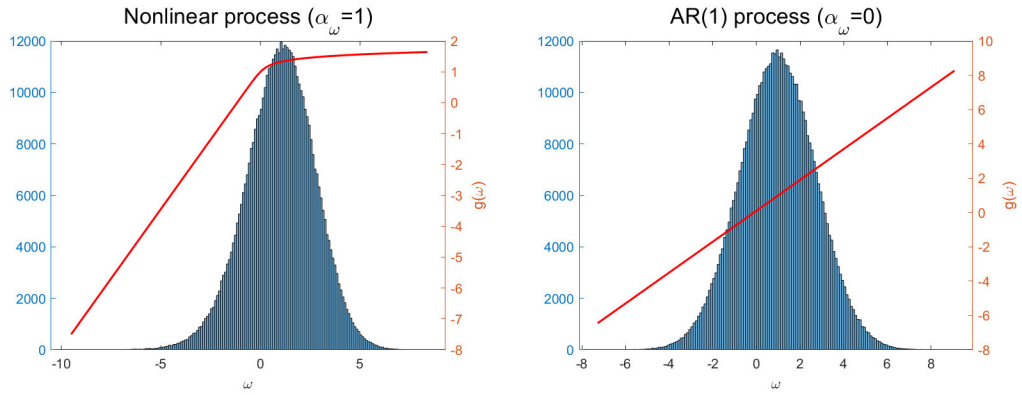


Figure 1: Distribution of productivity (left axis) and law of motion (right axis). Nonlinear process ($\alpha_\omega = 1$, left panel) and Gaussian $AR(1)$ process ($\alpha_\omega = 0$, right panel).

with q_{it}^*, q_{it}, and p_{it}. Note that because v_{it} is a function of k_{it}, p_{it}^V, ω_{it}, and δ_{it}, the firm's demand for the variable input cannot be inverted to express ω_{it} as a function of observables: invertibility fails as long as $\text{Var}(\delta_{1it}) > 0$ or $\text{Var}(\delta_{2it}) > 0$.

Recall that capital k_{it+1} is chosen in period t. We assume that capital fully depreciates between periods and endow the firm with static expectations regarding ω_{it+1}, δ_{it+1}, and p_{it+1}^V. Hence, the firm chooses k_{it+1} in period t given ω_{it}, δ_{it}, p_{it}^K, and p_{it}^V *as if* k_{it+1} and v_{it+1} are variable inputs.[14]

To complete the specification, we assume Gaussian $AR(1)$ processes for p_{it}^K, p_{it}^V, δ_{1it}, and δ_{2it}. We simulate $S = 100$ datasets with $N = 10,000$ firms and $T = 50$ periods. We provide further details on the specification in Appendix C.

Estimation. We estimate $\text{E}\left[q_{it}|x_{it}\right]$ in step 1 of the OP/LP/ACF procedure using the complete set of Hermite polynomials of total degree 4 in the variables in x_{it}. We estimate the parameters $\theta = (\alpha_0, \alpha_k, \rho, \nu, \mu_\omega, \rho_\omega, \alpha_\omega, \zeta_\omega)$ in step 2 using the complete set of Hermite polynomials of total degree 4 in the variables in z_{it} as instruments.[15] Our particular interest lies in comparing moment condition (5) with moment condition (12) to highlight the benefit of Neyman orthogonality in finite samples. We provide further details on the GMM estimator in Appendix C.

We proceed in three steps. First, we conduct a Lagrange multiplier test for the true parameter value θ^0. Our test corrects for the plug-in nature of the OP/LP/ACF procedure and clustering in

[14] Assuming constant returns to scale in production and quadratic adjustment costs to capital, the firm's dynamic programming problem that determines investment can be solved in closed form if the firm is a price-taker in the output market (Syverson 2001, Van Biesebroeck 2007, Ackerberg et al. 2015). As this is no longer possible under imperfect competition, we opt for a different, tractable specification of the evolution of capital.

[15] For a polynomial law of motion, it can be shown that a constant in the law of motion is not separately identified from a constant in the production function. While our law of motion is not polynomial, we nevertheless fix α_0 at its true value. If $\alpha_\omega = 0$, then we further fix ζ_ω at its true value.

the data. We provide further details on the Lagrange multiplier test in Appendix C.

Second, following the production function approach to markup estimation, we use the estimate of θ_f to estimate the markup $\mu_{it} = \frac{P_{it}}{MC_{it}}$ of firm i in period t as

$$\ln \mu_{it} + \varepsilon_{it} = p_{it} + q_{it} - p_{it}^V - v_{it} + \ln \frac{\partial f(k_{it}, v_{it})}{\partial v_{it}}.$$

The right-hand side is the log of the output elasticity minus the log of the expenditure share of the variable input.[16] Noting that the disturbance ε_{it} averages out as $\mathrm{E}[\varepsilon_{it}] = 0$, for simplicity we refer to the average of $\ln \mu_{it} + \varepsilon_{it}$ across firms and time as the average log markup.

Third, because the law of motion is of primary interest in some applications (Aw, Roberts & Xu 2011, De Loecker 2013, Doraszelski & Jaumandreu 2013), we use the estimate of θ_g to estimate

$$g'(0) = \rho_\omega \left((1 - \alpha_\omega) + \alpha_\omega \frac{1}{2 \ln 2} \right)$$

as a measure of persistence in the productivity process.

Results. We compare three cases for the observables x_{it} in step 1 of the OP/LP/ACF procedure and the instruments z_{it} in step 2:

- *Case 1:* $x_{it} = \left(k_{it}, v_{it}, p_{it}^V\right)$ and $z_{it} = \left(k_{it}, k_{it-1}, v_{it-1}, p_{it}^V, p_{it-1}^V\right)$.
- *Case 2:* $x_{it} = \left(k_{it+1}, k_{it}, v_{it}, p_{it+1}^V, p_{it}^V\right)$ and $z_{it} = \left(k_{it}, k_{it-1}, v_{it-1}, p_{it}^V, p_{it-1}^V\right)$.
- *Case 3:* $x_{it} = \left(k_{it+1}, k_{it}, v_{it}, p_{it}, p_{it+1}^V, p_{it}^V\right)$ and $z_{it} = \left(k_{it}, k_{it-1}, v_{it-1}, p_{it-1}, p_{it}^V, p_{it-1}^V\right)$.

Case 1 is our baseline and corresponds to how the OP/LP/ACF procedure is implemented in the existing literature.[17] Case 2 heeds Theorem 1 and ensures $x_{it-1} \supseteq z_{it}$. In light of Theorem 3, case 3 pursues a kitchen sink approach and adds the output price as a covariate that speaks to demand conditions.

Figure 2 shows the results of the OP/LP/ACF procedure with moment condition (5) for the nonlinear process ($\alpha_\omega = 1$). The Lagrange multiplier test for the true parameter value θ^0 rejects in all cases by a wide margin. In case 1, the failure of invertibility causes a massive bias in the production function and in the average log markup derived from it. Indeed, the average log markup

[16] De Loecker & Warzynski (2012) isolate $\ln \mu_{it}$ on the left-hand side and use the residual from the regression in step 1 of the OP/LP/ACF procedure to estimate ε_{it} on the right-hand side. However, using equation (6), the residual is $q_{it} - \mathrm{E}\left[q_{it} | x_{it}\right] = \varepsilon_{it} + \zeta_{it}$. Having ε_{it} on the left-hand side thus avoids that ζ_{it} taints the correlation of the estimated markup with a variable of interest such as the firm's export status or a measure of trade liberalization (Doraszelski & Jaumandreu 2024).

[17] We treat the price of the variable input p_{it}^V as observable to focus on demand shocks as the reason for the failure of invertibility. The nonidentification result in Gandhi, Navarro & Rivers (2020) assumes invertibility and therefore does not apply in our setting.

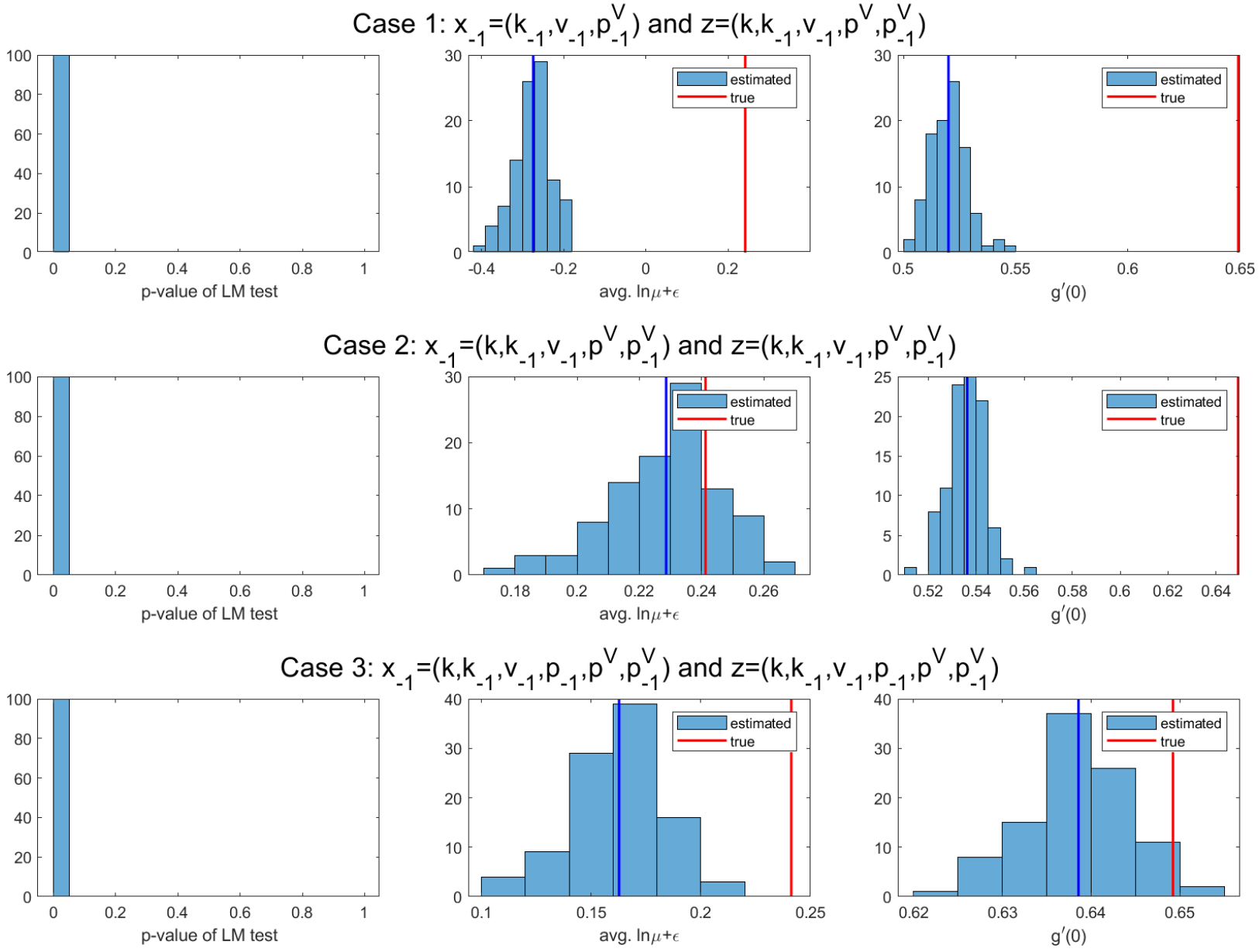


Figure 2: OP/LP/ACF procedure with moment condition (5). Nonlinear process ($\alpha_\omega = 1$). Case 1 (upper row), case 2 (middle row), and case 3 (lower row). p-value for Lagrange multiplier test (left column), average log markup (middle column), and measure of persistence $g'(0)$ (right column).

centers around -0.27 whereas short-run profit maximization implies $\ln \mu_{it} = \ln(1 + \exp(\delta_{2it})) > 0$ at the true parameter value θ^0. The failure of invertibility also causes a large bias in the law of motion as summarized by the measure of persistence $g'(0)$. In short, the failure of invertibility can have a substantial, detrimental impact on the OP/LP/ACF procedure as implemented in the existing literature.

In cases 2 and 3, ensuring $x_{it-1} \supseteq z_{it}$ in light of Theorem 1 greatly reduces the bias. The average log markup centers around 0.23 in case 2 and around 0.16 in case 3 compared to its true value of 0.24. In case 3, the measure of persistence $g'(0)$ further centers around 0.64 compared to its true value of 0.65.

Figure 3 shows the results of our modification the OP/LP/ACF procedure with moment condition (12) for the nonlinear process ($\alpha_\omega = 1$). Our modification reduces the bias in all cases. Case 3 combines our modification with adding the output price as a covariate that speaks to demand conditions. In this case, the average log markup centers around 0.25 compared to its true value of 0.24 and the measure of persistence $g'(0)$ centers around 0.64 compared to its true value of 0.65. The Lagrange multiplier test rejects for just 26 of the 100 simulated datasets at a 5% significance level.

Figure 4 shows the results for the OP/LP/ACF procedure with moment condition (5) for the Gaussian $AR(1)$ process ($\alpha_\omega = 0$). In case 1, the Lagrange multiplier test for the true parameter value θ^0 rejects by a wide margin and there are large biases. In cases 2 and 3, the Lagrange multiplier test does not reject. As in Example 1 in Section 4, ensuring $x_{it-1} \supseteq z_{it}$ eliminates the bias.

7 Sensitivity analysis

While setting $x_{it-1} \supseteq z_{it}$ provides a first-order bias correction, higher-order biases may remain and adversely affect the estimates of $\theta = (\theta_f, \theta_g)$ in step 2 of the OP/LP/ACF procedure. In this section, we develop a diagnostic to assess the sensitivity of the estimates to biases arising from the failure of invertibility. Our diagnostic remains valid without $x_{it-1} \supseteq z_{it}$.

To simplify the analysis, we abstract from any finite-samples differences between $e(x_{it-1})$ and $\mathrm{E}[q_{it-1}|x_{it-1}]$. We consider

$$\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(q^*_{it-1} - \lambda\zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)|z_{it}\right] = 0, \tag{13}$$

where $\lambda \geq 0$ scales the lagged prediction error ζ_{it-1}. If $\lambda = 0$, then moment condition (13) equals the valid but infeasible moment condition (4). If $\lambda = 1$, then moment condition (13) equals the original moment condition (5) because $\zeta_{it-1} = q^*_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]$ from equation (6). If, moreover, $x_{it-1} \supseteq z_{it}$, then it also equals the modified moment condition (12) with $e(x_{it-1}) = \mathrm{E}[q_{it-1}|x_{it-1}]$,

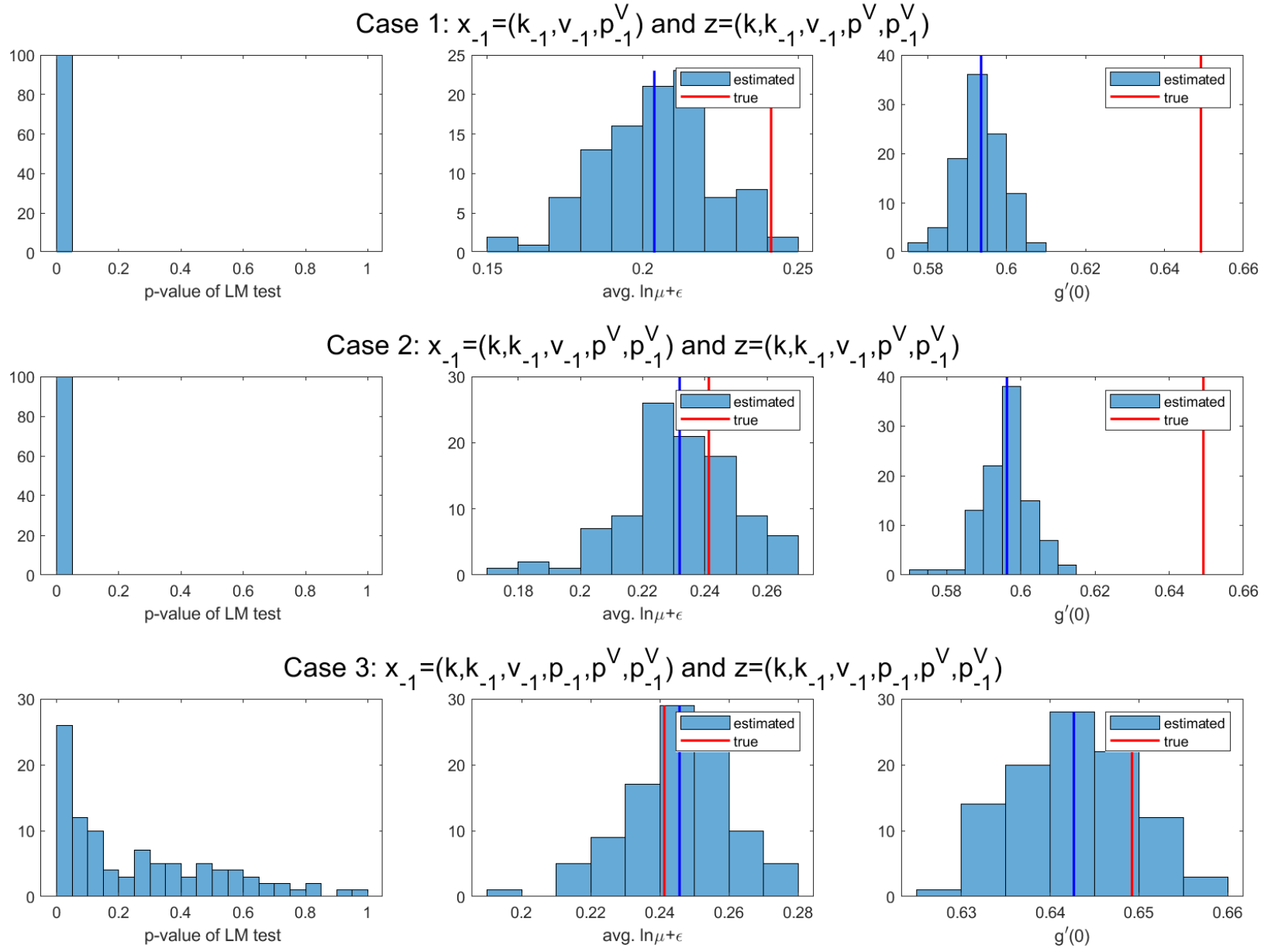


Figure 3: Modification of OP/LP/ACF procedure with moment condition (12). Nonlinear process ($\alpha_\omega = 1$). Case 1 (upper row), case 2 (middle row), and case 3 (lower row). p-value for Lagrange multiplier test (left column), average log markup (middle column), and measure of persistence $g'(0)$ (right column).

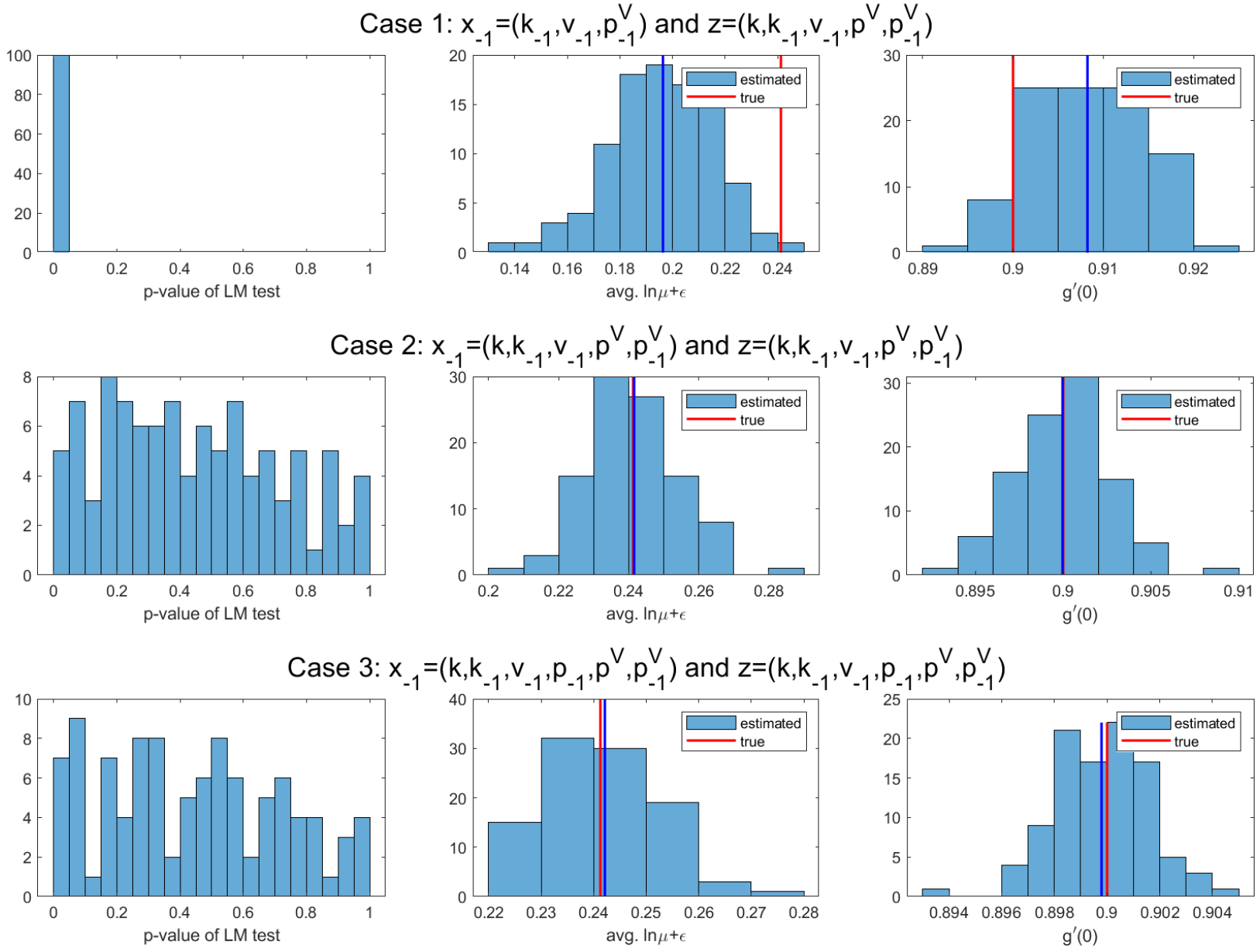


Figure 4: OP/LP/ACF procedure with moment condition (5). Gaussian $AR(1)$ process ($\alpha_\omega = 0$). Case 1 (upper row), case 2 (middle row), and case 3 (lower row). p-value for Lagrange multiplier test (left column), average log markup (middle column), and measure of persistence $g'(0)$ (right column).

as discussed in Section 5.

Define

$$m_{it}(\theta,\lambda) = q_{it} - f(k_{it}, v_{it}; \theta_f) - g\left(q^*_{it-1} - f(k_{it-1}, v_{it-1}; \theta_f) + \lambda\zeta_{it-1}; \theta_g\right),$$

where we make explicit the parameterization of the production function and law of motion. Moreover, define the pseudo-true value of θ as a function of λ as

$$\theta(\lambda) \in \arg\min_{\theta} \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]^\top W \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)],$$

where the superscript $\top$ denotes the transpose, $h(z_{it})$ is a vector function of the instruments z_{it}, and W is the GMM weighting matrix. By construction, $\theta(0)$ equals the true parameter value θ^0 and $\theta(1)$ is the probability limit of the GMM estimator based on either the original moment condition (5) or the modified moment condition (12). $\theta(1) - \theta(0)$ is therefore the bias.

Although it is not possible to consistently estimate the bias $\theta(1) - \theta(0)$, it is possible to consistently estimate $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ and thus to assess the sensitivity at the pseudo-true value $\theta(1)$. A small value of $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ provides assurance that small changes in the prediction error that arises in step 1 of the OP/LP/ACF procedure if invertibility fails do not dramatically alter the pseudo-true value $\theta(1)$. Indeed, one can show that $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ is zero in the examples in Section 4 where no bias arises despite the failure of invertibility. Conversely, a large value of $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ means that the pseudo-true value $\theta(1)$ is sensitive and thus warns of potentially large bias in the estimates.

To derive our diagnostic $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$, we show in Appendix E that the pseudo-true value $\theta(\lambda)$ under suitable regularity conditions satisfies the first-order condition

$$\mathrm{E}\left[\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\theta} h^\top(z_{it})\right] W \mathrm{E}[h(z_{it})m_{it}(\theta(\lambda),\lambda)] = 0 \tag{14}$$

where $\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\theta}$ is a $\dim(\theta)\times 1$ vector. The implicit function theorem implies

$$\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=1} = -\Gamma^{-1}\gamma, \tag{15}$$

where Γ is a $\dim(\theta)\times\dim(\theta)$ matrix and γ a $\dim(\theta)\times 1$ vector. Using $\frac{\partial m_{it}}{\partial\theta}$, $\frac{\partial m_{it}}{\partial\lambda}$, $\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top}$, and $\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda}$ to abbreviate $\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\theta}|_{\lambda=1}$, $\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\lambda}|_{\lambda=1}$, $\frac{\partial^2 m_{it}(\theta(\lambda),\lambda)}{\partial\theta\partial\theta^\top}|_{\lambda=1}$, and $\frac{\partial^2 m_{it}(\theta(\lambda),\lambda)}{\partial\theta\partial\lambda}|_{\lambda=1}$, respectively, the matrix Γ is defined as

$$\begin{aligned}\Gamma = &\left[\mathrm{E}\left[\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top} e_1 h^\top(z_{it})\right] W\mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)], \ldots, \mathrm{E}\left[\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top} e_{\dim(\theta)} h^\top(z_{it})\right] W\mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]\right] \\ &+ \mathrm{E}\left[\frac{\partial m_{it}}{\partial\theta} h^\top(z_{it})\right] W \mathrm{E}\left[h(z_{it})\frac{\partial m_{it}}{\partial\theta^\top}\right],\end{aligned} \tag{16}$$

where e_l is a $\dim(\theta) \times 1$ vector with a one in the lth position and zeros elsewhere. The vector γ is defined as

$$\gamma = \mathrm{E}\left[\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda}h^\top(z_{it})\right] W \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)] + \mathrm{E}\left[\frac{\partial m_{it}}{\partial\theta}h^\top(z_{it})\right] W \mathrm{E}\left[h(z_{it})\frac{\partial m_{it}}{\partial\lambda}\right]. \tag{17}$$

To evaluate our diagnostic $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$, we assume $\mathrm{E}[\varepsilon_{it-1}|x_{it-1}, z_{it}] = 0$. As we detail in Appendix E, consistently estimating Γ is straightforward. Consistently estimating γ is complicated by the fact that

$$\frac{\partial m_{it}}{\partial\lambda} = -\frac{\partial g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}}\zeta_{it-1}$$

and

$$\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda} = \begin{pmatrix} \frac{\partial^2 m_{it}}{\partial\theta_f\partial\lambda} \\ \frac{\partial^2 m_{it}}{\partial\theta_g\partial\lambda} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}^2}\frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}\zeta_{it-1} \\ -\frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}\partial\theta_g}\zeta_{it-1} \end{pmatrix}$$

depend on the lagged prediction error ζ_{it-1}. The key insight is that assuming $\mathrm{E}[\varepsilon_{it-1}|x_{it-1}, z_{it}] = 0$ implies

$$\mathrm{E}[\zeta_{it-1}|x_{it-1}, z_{it-1}] = \mathrm{E}[q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]\,|x_{it-1}, z_{it-1}].$$

Using the law of iterated expectations, we can substitute ζ_{it-1} with $q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]$ in the above expressions and rewrite γ as

$$\gamma = \mathrm{E}\left[\varphi_{it}h^\top(z_{it})\right] W \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)] + \mathrm{E}\left[\frac{\partial m_{it}}{\partial\theta}h^\top(z_{it})\right] W \mathrm{E}\left[h(z_{it})\psi_{it}\right], \tag{18}$$

where $\frac{\partial m_{it}}{\partial\lambda}$ and $\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda}$ are replaced by ψ_{it} and φ_{it} respectively with

$$\psi_{it} = -\frac{\partial g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}}(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}])$$

and

$$\varphi_{it} = \begin{pmatrix} \frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}^2}\frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]) \\ -\frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}\partial\theta_g}(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]) \end{pmatrix}.$$

We can therefore consistently estimate γ by the finite-sample analog to equation (18) using the estimate of $\mathrm{E}[q_{it-1}|x_{it-1}]$ from step 1 and that of θ from step 2.

Table 1 shows our diagnostic $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ for the nonlinear process ($\alpha_\omega = 1$). We evaluate the diagnostic at the theoretically best estimate of θ from our modification of the OP/LP/ACF procedure in cases 2 and 3. Because the above derivation presumes that moment condition (12) equals moment condition (5), we revert to the original procedure in case 1. We average the diagnostic over the $S = 100$ datasets from Section 6. For comparison, we also show the bias $\theta(1) - \theta(0)$.

In case 1, the large biases in the production function parameters $\theta_f = (\alpha_K, \rho, \nu)$ go hand-

	α_K	ρ	ν	μ_ω	ρ_ω	α_ω	ζ_ω
Case 1:							
Bias	0.1153	-0.3218	0.0075	0.1335	-0.0554	0.3778	-4.3693
Diagnostic	0.1482	-0.2849	0.0063	0.1962	-0.1378	0.1581	0.0752
	(0.0073)	(0.0501)	(0.0004)	(0.0656)	(0.0249)	(0.0849)	(0.3359)
Case 2:							
Bias	-0.0079	-0.0286	-0.0007	-0.1131	-0.0303	0.1277	-4.2358
Diagnostic	0.0109	-0.0114	-0.0013	-0.0651	-0.1124	-0.1169	-0.1361
	(0.0040)	(0.0266)	(0.0004)	(0.0102)	(0.0079)	(0.0151)	(0.0702)
Case 3:							
Bias	-0.0032	-0.0019	0.0003	-0.0209	-0.0060	0.0089	-0.6944
Diagnostic	-0.0018	-0.0728	0.0009	-0.0037	-0.0238	-0.0136	-0.0201
	(0.0010)	(0.0046)	(0.0001)	(0.0027)	(0.0021)	(0.0039)	(0.0919)

Table 1: Bias and diagnostic $\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\big|_{\lambda=1}$ averaged over $S = 100$ datasets with standard deviation in brackets. Nonlinear process ($\alpha_\omega = 1$). Case 1 uses estimate of θ based on moment condition (5) and cases 2 and 3 use estimate of θ based on moment condition (12).

in-hand with large values of the diagnostic. Conversely, in cases 2 and 3, the small biases go hand-in-hand with small values of the diagnostic. Turning to the law of motion parameters $\theta_g = (\mu_\omega, \rho_\omega, \alpha_\omega, \zeta_\omega)$, the relationship between bias and diagnostic is less tight. This may be because our specification of g is difficult to linearize around the sharp bend at the center of the distribution of productivity (see left panel of Figure 1). Nevertheless, the bias and the diagnostic are both an order of magnitude smaller in case 3 than in cases 1 and 2. Hence, the diagnostic remains useful to assess the sensitivity of the estimates to biases arising from the failure of invertibility.

Finally, we emphasize that our sensitivity analysis is conducted around the potentially misspecified model at $\lambda = 1$ as opposed to the true model at $\lambda = 0$ and remains valid regardless of the magnitude of misspecification. Our approach thus differs from the sensitivity analysis in Andrews et al. (2017) that is only valid locally around the true model.

8 Concluding remarks

We revisit the OP/LP/ACF procedure for production function estimation by scrutinizing the invertibility assumption at its core. We show that this assumption is testable and strongly rejected in widely used panel data. A failure of invertibility has important consequences: the prediction of planned output q_{it}^* from observables x_{it} in step 1 of the OP/LP/ACF procedure contains an error that invalidates capital k_{it} as an instrument, leading to biased estimates in step 2.

Fortunately, much can still be done. We establish a necessary and sufficient condition for

the moment condition in step 2 of the OP/LP/ACF procedure to hold for the true production function. This condition compels a rethinking of the OP/LP/ACF procedure: any instrument used in step 2 must be appropriately included in the regression in step 1 to ensure $x_{it-1} \supseteq z_{it}$. At a minimum, this calls for adding the lead of capital k_{it+1} to x_{it}. Furthermore, our condition is easier to satisfy if the law of motion is modelled flexibly and we show that taking a kitchen sink approach to the regression in step 1 mitigates violations of our condition.

In case our condition is violated, we show that setting $x_{it-1} \supseteq z_{it}$ provides a first-order bias correction. Explicitly integrating the bias correction achieves Neyman orthogonality in the modified moment condition. Both theoretical arguments and Monte Carlo simulations underscore that this modification can substantially improve the finite-sample performance of the GMM estimator in step 2, especially if the regression in step 1 uses a rich set of covariates.

Finally, we recognize that, despite ensuring $x_{it-1} \supseteq z_{it}$, higher-order biases may remain. To gauge their importance for the GMM estimator in step 2, we introduce a diagnostic that measures the sensitivity of the estimates to the size of the prediction error that arises in step 1 from a failure of invertibility.

In sum, the invertibility assumption is demanding and can fail because of unobserved demand heterogeneity or in imperfectly competitive environments with partially or completely unobserved rivals or changes in firm conduct and for other reasons. It is therefore particularly challenging for the production function approach to markup estimation. By outlining how to test the invertibility assumption and offering straightforward modifications of the OP/LP/ACF procedure that either eliminate or mitigate the bias that results from a failure of invertibility, we provide a practical framework for production function estimation under broader conditions. We hope that Neyman orthogonality, together with our diagnostic, are valuable tools for researcher seeking to improve finite-sample performance and assessing the robustness of their empirical findings.

Appendix

A Tests for invertibility

Tests. Delgado & Manteiga (2001) and Cai, Guo & Zhong (2024) develop nonparametric inference procedures for a mean-independence restriction (Proposition 1), while Shah & Peters (2020) focus on a conditional-independence restriction (Proposition 2). These procedures presume i.i.d. data. Instead of adapting them for panel data, we proceed semiparametrically in what follows.

We test $\mathrm{E}[q_{it}|x_{it}, x_{it-1}] = \mathrm{E}[q_{it}|x_{it}]$ in Proposition 1 using the partially linear regression model

$$q_{it} = x_{it-1}^{\top}\beta + \psi(x_{it}) + e_{it}, \tag{19}$$

where the superscript $\top$ denotes the transpose and $\psi(x_{it})$ is a sufficiently flexible function. In practice, we specify $\psi(x_{it}) = r^{\top}(x_{it})\gamma$, where $r(x_{it})$ is the complete set of polynomials of a given total degree in the variables in x_{it} and γ is a parameter vector. $\mathrm{E}[q_{it}|x_{it}, x_{it-1}] = \mathrm{E}[q_{it}|x_{it}]$ implies $\beta = 0$. Rejecting $H_0 : \beta = 0$ is therefore evidence against invertibility. To implement this test, we use OLS to estimate the partially linear model (19) and conduct a standard F-test with firm-level clustering.

Because the partially linear model (19) involves a large number of regressors, the standard F-test may perform poorly in finite samples. To address this concern, we use the procedure in Chernozhukov, Hansen & Spindler (2015) (henceforth, CHS) instead of OLS. This allows us to use LASSO for dimensionality reduction while achieving oracle efficiency for the estimator of β via Neyman orthogonalization.

We test $q_{it} \perp\!\!\!\perp x_{it-1}|x_{it}$ in Proposition 2 using quantile regression to estimate the partially linear model (19). Let $Q_\tau(q_{it}|x_{it})$ and $Q_\tau(q_{it}|x_{it}, x_{it-1})$ denote the τth quantile of q_{it} conditional on x_{it} and (x_{it}, x_{it-1}), respectively. $q_{it} \perp\!\!\!\perp x_{it-1}|x_{it}$ implies $\beta = 0$ and is equivalent to $Q_\tau(q_{it}|x_{it}) = Q_\tau(q_{it}|x_{it}, x_{it-1})$ for all $\tau \in (0, 1)$. We restrict attention to the median ($\tau = 0.5$) and conduct the F-test with firm-level clustering in Parente & Santos Silva (2016).

Data. We implement our tests on two widely used datasets: an unbalanced panel of Spanish manufacturing firms from 1990 to 2006 (Encuesta Sobre Estrategias Empresariales, henceforth ESEE) and the balanced panel of US manufacturing industries from 1958 to 2018 (NBER-CES) from Becker, Gray & Marvakov (2021).

We follow the data cleaning procedures and variable definitions in Doraszelski & Jaumandreu (2018, 2024) for the ESEE data and in Jaumandreu (2022) and Jaumandreu & Lopez (2024) for the NBER-CES data. The available variables include output q_{it}, capital k_{it}, labor l_{it}, materials m_{it}, the price of output p_{it}, the wage w_{it}, and the price of materials p_{it}^M (in logs). In addition, the

			total degree 2			total degree 3		total degree 4	
industry	#obs	#firms	OLS	CHS	Median	OLS	CHS	OLS	CHS
1	2365	296	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1270	147	0.004	0.043	0.046	0.008	0.032	0.012	0.033
3	2168	278	0.000	0.000	0.002	0.000	0.000	0.000	0.000
4	1411	165	0.000	0.004	0.005	0.016	0.006	0.035	0.004
5	1505	197	0.000	0.004	0.012	0.000	0.000	0.000	0.000
6	1206	148	0.000	0.000	0.018	0.001	0.000	0.001	0.000
7	2455	304	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	2368	317	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	1445	191	0.001	0.000	0.001	0.017	0.000	0.009	0.000
10	1414	174	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 2: p-value for test of $H_0 : \beta = 0$ with $x_{it} = (k_{it}, l_{it}, m_{it}, p_{it}, w_{it}, p_{it}^M)$. ESEE data.

			total degree 2			total degree 3		total degree 4	
industry	#obs	#firms	OLS	CHS	Median	OLS	CHS	OLS	CHS
1	2365	296	0.000	0.000	0.000	0.000	0.000	0.000	0.000
2	1270	147	0.000	0.001	0.110	0.001	0.000	0.070	0.001
3	2168	278	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	1411	165	0.001	0.000	0.107	0.062	0.000	0.208	0.000
5	1505	197	0.000	0.000	0.004	0.000	0.000	0.000	0.000
6	1206	148	0.000	0.000	0.017	0.000	0.000	0.015	0.000
7	2455	304	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	2368	317	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	1445	191	0.000	0.000	0.009	0.000	0.000	0.007	0.000
10	1414	174	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 3: p-value for test of $H_0 : \beta = 0$ with $x_{it} = (k_{it}, l_{it}, m_{it}, p_{it}, w_{it}, p_{it}^M, mdy_{it})$. ESEE data.

ESEE data has a market dynamism variable mdy_{it} that indicates shifts in the demand the firm faces (slump, stability, or expansion). Investment is non-zero in the NBER-CES data, so that the investment variable inv_{it} (in logs) is well-defined.

Results. Tables 2 and 3 present results for the 10 industries in the ESEE data. We show the p-value for the test of $H_0 : \beta = 0$ for different estimators (OLS, CHS, and median regression) and for different specifications of $g(x_{it})$ (total degree 2, 3, and 4).[18] In Table 2, we use $x_{it} = (k_{it}, l_{it}, m_{it}, p_{it}, w_{it}, p_{it}^M)$. We reject invertibility by a wide margin in all industries and for all combinations of estimators and specifications of $g(x_{it})$. In Table 3, we add the market dynamism

[18]Results for median regression are limited to total degree 2 because the Stata package `qreg2` fails to converge for total degree 3 and 4.

variable mdy_{it} to x_{it}. While we continue to reject invertibility in 8 industries, we cannot reject invertibility at a 5% significance level in industries 2 and 4 for some combinations of estimators and specifications of $g(x_{it})$. This reinforces that invertibility hinges on the ability to control for demand.

We note that all estimated models in Tables 2 and 3 have an R^2 of at least 0.990 and some have an R^2 of 0.998. Insufficient flexibility in modeling $g(x_{it})$ is thus not likely to compromise our results.

Turning to the NBER-CES data, we alternatively use $x_{it} = (k_{it}, l_{it}, m_{it}, p_{it}, w_{it}, p_{it}^M)$ and $x_{it} = (k_{it}, l_{it}, m_{it}, p_{it}, w_{it}, p_{it}^M, inv_{it})$. The latter follows OP to invert the firm's demand for investment. Throughout the p-value for the test of $H_0 : \beta = 0$ is 0.000 and the R^2 exceeds 0.987.

B Failure of invertibility and validity of OP/LP/ACF moment condition

Omitted proofs.

Proof of Theorem 2. Using $\zeta_{it-1} = q_{it-1}^* - \mathrm{E}[q_{it-1}|x_{it-1}]$, we have

$$\begin{aligned}
& \mathrm{E}\left[g'\left(q_{it-1}^* - \zeta_{it-1} - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1}\middle|z_{it}\right] \\
= \; & \mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1}\middle|z_{it}\right] \\
= \; & \mathrm{E}\left[\mathrm{E}[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\zeta_{it-1}|x_{it-1}, z_{it}]\middle|z_{it}\right] \\
= \; & \mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\mathrm{E}[\zeta_{it-1}|x_{it-1}, z_{it}]\middle|z_{it}\right] \\
= \; & 0, \qquad (20)
\end{aligned}$$

where the second equality is due to the law of iterated expectations, the third equality uses that $x_{it-1} = (k_{it-1}, v_{it-1}, \ldots)$ implies that $g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)$ is a function of x_{it-1}, and the last equality uses that $x_{it-1} \supseteq z_{it}$ implies $\mathrm{E}[\zeta_{it-1}|x_{it-1}, z_{it}] = \mathrm{E}[\zeta_{it-1}|x_{it-1}] = 0$. Finally, using equation (20) and $\zeta_{it-1} = q_{it-1}^* - \mathrm{E}[q_{it-1}|x_{it-1}]$ establishes equation (10). □

Proof of Theorem 3. Because $g^0 \in \mathcal{G}$, we have

$$\begin{aligned}
& \inf_{\tilde{g}\in\mathcal{G}} \left\|\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] - \mathrm{E}\left[\tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)|z_{it}\right]\right\|_{L,1} \\
= \; & \inf_{\tilde{g}\in\mathcal{G}} \mathrm{E}\left[\left|\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] - \mathrm{E}\left[\tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)|z_{it}\right]\right|\right] \\
\leq \; & \mathrm{E}\left[\left|\mathrm{E}[g^0(\omega_{it-1})|z_{it}] - \mathrm{E}[g^0(\mathrm{E}[\omega_{it-1}|x_{it-1}])|z_{it}]\right|\right].
\end{aligned}$$

Using a second-order Taylor expansion of g^0 around $\mathrm{E}[\omega_{it-1}|x_{it-1}]$, we have

$$\begin{aligned}
& \left|\mathrm{E}[g^0(\omega_{it-1})|z_{it}] - \mathrm{E}[g^0(\mathrm{E}[\omega_{it-1}|x_{it-1}])|z_{it}]\right| \\
= & \left|\mathrm{E}\left[g^{0\prime}(\mathrm{E}[\omega_{it-1}|x_{it-1}])\zeta_{it-1}\middle|z_{it}\right] + \mathrm{E}\left[g^{0\prime\prime}(\tilde{\omega}_{it-1})\zeta_{it-1}^2\middle|z_{it}\right]\right| \\
\leq & \left|\mathrm{E}\left[g^{0\prime}(\mathrm{E}[\omega_{it-1}|x_{it-1}])\zeta_{it-1}\middle|z_{it}\right]\right| + \left|\mathrm{E}\left[g^{0\prime\prime}(\tilde{\omega}_{it-1})\zeta_{it-1}^2\middle|z_{it}\right]\right| \\
\leq & \left|\mathrm{E}\left[g^{0\prime}(\mathrm{E}[\omega_{it-1}|x_{it-1}])\zeta_{it-1}\middle|z_{it}\right]\right| + \tau\mathrm{E}\left[\zeta_{it-1}^2\middle|z_{it}\right],
\end{aligned}$$

where $\tilde{\omega}_{it-1}$ is some value between ω_{it-1} and $\mathrm{E}[\omega_{it-1}|x_{it-1}]$, and the last inequality follows from the definition of τ. As in the proof of Theorem 2, because $x_{it-1} \supseteq z_{it}$, the law of iterated expectations implies $\mathrm{E}\left[g^{0\prime}(\mathrm{E}[\omega_{it-1}|x_{it-1}])\zeta_{it-1}\middle|z_{it}\right] = 0$. Thus, we have

$$\inf_{\tilde{g}\in\mathcal{G}} \left\|\mathrm{E}\left[g^0(\omega_{it-1})|z_{it}\right] - \mathrm{E}\left[\tilde{g}\left(\mathrm{E}\left[\omega_{it-1}|x_{it-1}\right]\right)|z_{it}\right]\right\|_{L,1} \leq \mathrm{E}\left[\tau\mathrm{E}\left[\zeta_{it-1}^2|z_{it}\right]\right] = \tau\mathrm{Var}(\zeta_{it-1}).$$

□

C Monte Carlo exercise

Data generating process. Marginal revenue is

$$mr_{it} = -\ln\left(1+\exp(\delta_{2it})\right) + p_{it}$$

and marginal cost is

$$mc_{it} = p_{it}^V + v_{it} - f(k_{it}, v_{it}) - \ln\frac{\partial f(k_{it}, v_{it})}{\partial v_{it}} - \omega_{it}.$$

We rewrite $mr_{it} = mc_{it}$ as

$$\frac{\delta_{1it} + \exp(-\delta_{2it})(f(k_{it}, v_{it}) + \omega_{it})}{1+\exp(-\delta_{2it})} - \ln\left(1+\exp(\delta_{2it})\right) + \ln\frac{\partial f(k_{it}, v_{it})}{\partial v_{it}} - p_{it}^V - v_{it} = 0$$

and solve numerically for v_{it}. With v_{it} in hand, we determine $q_{it}^* = f(k_{it}, v_{it}) + \omega_{it}$, $q_{it} = q_{it}^* + \varepsilon_{it}$, and $p_{it} = \frac{\delta_{1it} - q_{it}^*}{1+\exp(-\delta_{2it})}$.

As capital fully depreciates between periods and the firm has static expectations, it chooses k_{it+1} in period t given ω_{it}, δ_{it}, p_{it}^K, and p_{it}^V whilst anticipating that it optimally chooses $\tilde{v}_{it+1}$ in

period $t+1$. To maximize its expected profit in period $t+1$, in period t the firm solves

$$\begin{aligned}&\max_{k_{it+1}}\left(\max_{\tilde{v}_{it+1}}\exp(\tilde{p}_{it+1}+\tilde{q}^*_{it+1})-\exp(p^K_{it}+k_{it+1})-\exp(p^V_{it}+\tilde{v}_{it+1})\right)\\&=\max_{k_{it+1},\tilde{v}_{it+1}}\exp(\tilde{p}_{it+1}+\tilde{q}^*_{it+1})-\exp(p^K_{it}+k_{it+1})-\exp(p^V_{it}+\tilde{v}_{it+1})\end{aligned}$$

subject to

$$\begin{aligned}\tilde{q}^*_{it+1}&=f(k_{it+1},\tilde{v}_{it+1})+\omega_{it},\\\tilde{q}^*_{it+1}&=\delta_{1it}-(1+\exp(-\delta_{2it}))\tilde{p}_{it+1}.\end{aligned}$$

The solution is

$$\begin{aligned}k_{it+1}=\frac{1}{1-\rho}\left(\ln\alpha_K-p^K_{it}\right)+\frac{\exp(\delta_{2it})+1}{\exp(\delta_{2it})+1-\nu}\Bigg(&\ln\nu\\+\left(\frac{\nu}{(\exp(\delta_{2it})+1)\rho}-1\right)\ln\left(\alpha_K\left(\frac{\alpha_K}{\exp(p^K_{it})}\right)^{\frac{\rho}{1-\rho}}+(1-\alpha_K)\left(\frac{1-\alpha_K}{\exp(p^V_{it})}\right)^{\frac{\rho}{1-\rho}}\right)\\-\ln\left(\exp(\delta_{2it})+1\right)+\frac{1}{\exp(\delta_{2it})+1}\left(\alpha_0+\omega_{it}\right)+\frac{\delta_{1it}}{1+\exp(-\delta_{2it})}\Bigg).\end{aligned}$$

GMM estimator. Corresponding to moment condition (5) in step 2, define the moment function

$$m_{it}(\theta)=q_{it}-f(k_{it},v_{it};\theta_f)-g\left(e(x_{it-1})-f(k_{it-1},v_{it-1};\theta_f);\theta_g\right), \tag{21}$$

where we replace $e(x_{it-1})$ by the estimate from step 1. In step 2, we solve the GMM problem

$$\min_\theta\left(\frac{1}{NT}\sum_{i,t}h(z_{it})m_{it}(\theta)\right)^\top W\left(\frac{1}{NT}\sum_{i,t}h(z_{it})m_{it}(\theta)\right), \tag{22}$$

where the superscript $\top$ denotes the transpose, $h(z_{it})$ is the complete set of Hermite polynomials of total degree 4 in the variables in z_{it}, and

$$W=\left(\frac{1}{NT-1}\sum_{i,t}\left(h(z_{it})m_{it}\left(\theta^0\right)-\hat{\mu}\right)^\top\left(h(z_{it})m_{it}\left(\theta^0\right)-\hat{\mu}\right)\right)^{-1}\text{ with }\hat{\mu}=\frac{1}{NT}\sum_{i,t}h(z_{it})m_{it}\left(\theta^0\right),$$

is a weighting matrix evaluated at the true parameters values θ^0.

Corresponding to moment condition (12), our modification of the OP/LP/ACF procedure

redefines the moment function as

$$m_{it}(\theta) = q_{it} - f(k_{it}, v_{it}; \theta_f) - g\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1}; \theta_f); \theta_g\right) \\ - g'\left(e(x_{it-1}) - f(k_{it-1}, v_{it-1}; \theta_f); \theta_g\right)\left(q_{it-1} - e(x_{it-1})\right), \quad (23)$$

where we replace $e(x_{it-1})$ by the estimate from step 1. We hold fixed W to facilitate the comparison between our modification and the original procedure.

Lagrange multiplier (LM) test. We first derive the LM test for the original moment function (21). To account for the plug-in nature of the GMM estimator in step 2, we have to explicitly specify how $e(x_{it-1})$ is modelled and estimated in step 1. In our Monte Carlo exercise, we model $e(x_{it-1}) = r^\top(x_{it-1})\tau$, where the superscript $\top$ denotes the transpose, $r(x_{it-1})$ is the complete set of Hermite polynomials of total degree 4 in the variables in x_{it-1}, and τ is a parameter vector. Consequently, we define

$$m_{it}(\theta, \tau) = q_{it} - f(k_{it}, v_{it}; \theta_f) - g\left(r^\top(x_{it-1})\tau - f(k_{it-1}, v_{it-1}; \theta_f); \theta_g\right), \quad (24)$$

where we make explicit the dependence of the moment function on τ. We estimate τ in step 1 by OLS as $\hat{\tau} = [\frac{1}{NT}\sum_{i,t} r(x_{it-1})r^\top(x_{it-1})]^{-1}\frac{1}{NT}\sum_{i,t} q_{it-1}r(x_{it-1})$.[19]

Following standard asymptotic arguments, we can show that

$$\widehat{\Sigma}^{-\frac{1}{2}}\frac{1}{\sqrt{NT}}\sum_{i,t} h(z_{it})m_{it}(\theta^0, \hat{\tau}) \xrightarrow{d} N(0, I),$$

where $\widehat{\Sigma} = \widehat{\Lambda}\widehat{\Omega}\widehat{\Lambda}^\top$,

$$\widehat{\Lambda} = \left(I \quad \sum_{i,t} h(z_{it})\left[\frac{\partial m_{it}(\theta^0, \hat{\tau})}{\partial \tau^\top}\right]\left[\sum_{i,t} r(x_{it-1})r^\top(x_{it-1})\right]^{-1}\right),$$

[19] We can alternatively estimate τ as $\hat{\tau} = [\frac{1}{NT}\sum_{i,t} r(x_{it})r^\top(x_{it})]^{-1}\frac{1}{NT}\sum_{i,t} q_{it}r(x_{it})$. In our simulated datasets, both current and lagged values are available for all T periods and N firms.

with I being the $\dim(h(z_{it})) \times \dim(h(z_{it}))$ identity matrix, and

$$\begin{aligned}
\widehat{\Omega} &= \begin{pmatrix} \widehat{\Omega}_{11} & \widehat{\Omega}_{12} \\ \widehat{\Omega}_{12}^{\top} & \widehat{\Omega}_{22} \end{pmatrix}, \\
\widehat{\Omega}_{11} &= \frac{1}{NT} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{i=1}^{N} \left[m_{it}(\theta^0, \hat{\tau}) m_{it'}(\theta^0, \hat{\tau}) h(z_{it}) h^{\top}(z_{it'}) \right], \\
\widehat{\Omega}_{12} &= \frac{1}{NT} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{i=1}^{N} \left[m_{it}(\theta^0, \hat{\tau}) (q_{it'-1} - r^{\top}(x_{it'-1})\hat{\tau}) h(z_{it}) r^{\top}(x_{it'-1}) \right], \\
\widehat{\Omega}_{22} &= \frac{1}{NT} \sum_{t=1}^{T} \sum_{t'=1}^{T} \sum_{i=1}^{N} \left[(q_{it-1} - r^{\top}(x_{it-1})\hat{\tau}) (q_{it'-1} - r^{\top}(x_{it'-1})\hat{\tau}) r(x_{it-1}) r^{\top}(x_{it'-1}) \right].
\end{aligned}$$

Note that $\widehat{\Sigma}$ incorporates firm-level clustering and thus allows for arbitrary correlation within firm across time in the data. As $N \to \infty$, the LM statistic

$$S_N = \frac{1}{NT} \left(\sum_{i,t} h(z_{it}) m_{it}(\theta^0, \hat{\tau}) \right)^{\top} \widehat{\Sigma}^{-1} \left(\sum_{i,t} h(z_{it}) m_{it}(\theta^0, \hat{\tau}) \right)$$

converges to a χ^2 distribution with $\dim(h(z_{it}))$ degrees of freedom. Let $c_{1-\alpha}$ be the $1-\alpha$ quantile of this distribution. Then rejecting $H_0 : \mathrm{E}[h(z_{it}) m_{it}(\theta^0, \tau^0)] = 0$ if $S_N > c_{1-\alpha}$ is a valid LM inference procedure with asymptotic size α.

Turning to the modified moment function (23), Neyman orthogonality renders correcting for the plug-in nature of the GMM estimator in step 2 superfluous. Consequently, we use a standard LM test with firm-level clustering. As $N \to \infty$, the LM statistic

$$S_N = \frac{1}{NT} \left(\sum_{i,t} h(z_{it}) m_{it}(\theta^0, \hat{\tau}) \right)^{\top} \widehat{\Omega}_{11}^{-1} \left(\sum_{i,t} h(z_{it}) m_{it}(\theta^0, \hat{\tau}) \right)$$

converges to a χ^2 distribution with $\dim(h(z_{it}))$ degrees of freedom. Rejecting $H_0 : \mathrm{E}[h(z_{it}) m_{it}(\theta^0, \tau^0)] = 0$ if $S_N > c_{1-\alpha}$ is a valid LM inference procedure with asymptotic size α.

D Modification of OP/LP/ACF moment condition

Omitted proofs.

Proof of Theorem 4. By the definition of the derivative, it suffices to show that for any non-zero

sequence λ_n with $\lim_{n\to\infty}\lambda_n = 0$, we have

$$\lim_{n\to\infty}\frac{1}{\lambda_n}\Bigg[\Bigg(\mathrm{E}\left[q_{it} - f(k_{it}, v_{it}) - g\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\right.$$
$$\left. - g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}] - \lambda_n\delta(x_{it-1})\right)\middle|z_{it}\right]\Bigg)$$
$$- \Bigg(\mathrm{E}\Big[q_{it} - f(k_{it}, v_{it}) - g\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)$$
$$\left. - g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]\right)\middle|z_{it}\Big]\right)\Bigg] = 0. \quad (25)$$

Similar to the idea used in the proof of Theorem 2, the law of iterated expectations implies

$$\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]\right)|z_{it}\right] = 0, \quad (26)$$
$$\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\left(q_{it-1} - \mathrm{E}[q_{it-1}|x_{it-1}]\right)|z_{it}\right] = 0. \quad (27)$$

Using equations (26) and (27), equation (25) is equivalent to

$$\lim_{n\to\infty}\frac{1}{\lambda_n}\mathrm{E}\left[g\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right) - g\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\middle|z_{it}\right]$$
$$= \lim_{n\to\infty}\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\delta(x_{it-1})\middle|z_{it}\right]. \quad (28)$$

It therefore remains to establish equation (28). We do this by showing that both sides of equation (28) are equal to $\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\delta(x_{it-1})\middle|z_{it}\right]$.

Because g' is bounded and δ is integrable, there exists some $\eta > 0$ such that

$$\mathrm{E}\left[\sup_{\lambda:|\lambda|<\eta}\left|g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\delta(x_{it-1})\right|\middle|z_{it}\right] < \infty \quad (29)$$

almost surely.[20] Because g' is continuous, the dominated convergence theorem therefore implies that

$$\lim_{n\to\infty}\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] + \lambda_n\delta(x_{it-1}) - f(k_{it-1}, v_{it-1})\right)\delta(x_{it-1})\middle|z_{it}\right]$$
$$= \mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}] - f(k_{it-1}, v_{it-1})\right)\delta(x_{it-1})\middle|z_{it}\right]. \quad (30)$$

[20] Theorem 4 continues to hold if the assumption that g' is bounded is relaxed to this condition.

Turning to the other side of equation (28), the mean value theorem implies that

$$\begin{aligned}&\frac{1}{\lambda_n}\left(g\left(\mathrm{E}[q_{it-1}|x_{it-1}]+\lambda_n\delta(x_{it-1})-f(k_{it-1},v_{it-1})\right)-g\left(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1})\right)\right)\\ =&g'\left(\mathrm{E}[q_{it-1}|x_{it-1}]+\tilde{\lambda}_n\delta(x_{it-1})-f(k_{it-1},v_{it-1})\right)\delta(x_{it-1}),\end{aligned} \tag{31}$$

where $\tilde{\lambda}_n$ is some value between 0 and λ_n that implicitly depends on x_{it-1}. Because $|\tilde{\lambda}_n| \leq |\lambda_n|$ and g' is continuous, we have

$$\begin{aligned}&\lim_{n\to\infty} g'\left(\mathrm{E}[q_{it-1}|x_{it-1}]+\tilde{\lambda}_n\delta(x_{it-1})-f(k_{it-1},v_{it-1})\right)\delta(x_{it-1})\\ =\;& g'\left(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1})\right)\delta(x_{it-1})\end{aligned} \tag{32}$$

for all x_{it-1}. Given inequality (29) and equation (32), the dominated convergence theorem implies that

$$\begin{aligned}&\lim_{n\to\infty}\frac{1}{\lambda_n}\mathrm{E}\left[g\left(\mathrm{E}[q_{it-1}|x_{it-1}]+\lambda_n\delta(x_{it-1})-f(k_{it-1},v_{it-1})\right)-g\left(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1})\right)\Big|z_{it}\right]\\ =\;&\mathrm{E}\left[g'\left(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1})\right)\delta(x_{it-1})\Big|z_{it}\right].\end{aligned}$$

This establishes equation (28). □

E Sensitivity analysis

Assume $f(k_{it},v_{it};\theta_f)$ is twice continuously differentiable in θ_f and $g(\omega_{it-1};\theta_g)$ is twice continuously differentiable in (θ_g,ω_{it-1}), so that $m_{it}(\theta,\lambda)$ is twice continuously differentiable in (θ,λ). Assume the product of h and the first-order derivatives of $m_{it}(\theta,\lambda)$ is always integrable. Then the dominated convergence theorem implies

$$\frac{\partial \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]}{\partial\theta^\top}=\mathrm{E}\left[h(z_{it})\frac{\partial m_{it}(\theta,\lambda)}{\partial\theta^\top}\right],\quad \frac{\partial \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]}{\partial\lambda}=\mathrm{E}\left[h(z_{it})\frac{\partial m_{it}(\theta,\lambda)}{\partial\lambda}\right].$$

Similarly, assume the product of h and the second-order derivatives of $m_{it}(\theta,\lambda)$ is always integrable. Then the dominated convergence theorem implies

$$\frac{\partial^2 \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]}{\partial\theta\partial\theta^\top}=\mathrm{E}\left[h(z_{it})\frac{\partial^2 m_{it}(\theta,\lambda)}{\partial\theta\partial\theta^\top}\right],\quad \frac{\partial^2 \mathrm{E}[h(z_{it})m_{it}(\theta,\lambda)]}{\partial\theta\partial\lambda}=\mathrm{E}\left[h(z_{it})\frac{\partial^2 m_{it}(\theta,\lambda)}{\partial\theta\partial\lambda}\right].$$

Under these regularity conditions, the pseudo-true value $\theta(\lambda)$ satisfies the first-order condition (14).

Assume the first-order condition (14) identifies $\theta(\lambda)$ for λ in a neighborhood of 1. Then the

implicit function theorem implies

$$\mathrm{E}\left[\left(\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top}\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=1}+\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda}\right)h^\top(z_{it})\right]W\mathrm{E}\left[h(z_{it})m_{it}(\theta(\lambda),\lambda)\right]$$
$$+\mathrm{E}\left[\frac{\partial m_{it}}{\partial\theta}h^\top(z_{it})\right]W\mathrm{E}\left[h(z_{it})\left(\frac{\partial m_{it}}{\partial\theta^\top}\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=1}+\frac{\partial m_{it}}{\partial\lambda}\right)\right]=0, \quad (33)$$

where we use $\frac{\partial m_{it}}{\partial\theta}$, $\frac{\partial m_{it}}{\partial\lambda}$, $\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top}$, and $\frac{\partial^2 m_{it}}{\partial\theta\partial\lambda}$ to abbreviate $\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\theta}|_{\lambda=1}$, $\frac{\partial m_{it}(\theta(\lambda),\lambda)}{\partial\lambda}|_{\lambda=1}$, $\frac{\partial^2 m_{it}(\theta(\lambda),\lambda)}{\partial\theta\partial\theta^\top}|_{\lambda=1}$, and $\frac{\partial^2 m_{it}(\theta(\lambda),\lambda)}{\partial\theta\partial\lambda}|_{\lambda=1}$, respectively. Equation (33) can be rewritten as

$$\Gamma\frac{\mathrm{d}\theta(\lambda)}{\mathrm{d}\lambda}\Big|_{\lambda=1}+\gamma=0,$$

where Γ and γ are defined in equations (16) and (17), respectively. Assume Γ is invertible. Then equation (15) follows.

Consistently estimating Γ is straightforward because

$$\frac{\partial m_{it}}{\partial\theta}=\begin{pmatrix}\frac{\partial m_{it}}{\partial\theta_f}\\ \frac{\partial m_{it}}{\partial\theta_g}\end{pmatrix}=\begin{pmatrix}-\frac{\partial f(k_{it},v_{it};\theta_f)}{\partial\theta_f}+\frac{\partial g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}}\frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}\\ -\frac{\partial g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\theta_g}\end{pmatrix}$$

and

$$\frac{\partial^2 m_{it}}{\partial\theta\partial\theta^\top}=\begin{pmatrix}\frac{\partial^2 m_{it}}{\partial\theta_f\partial\theta_f^\top} & \frac{\partial^2 m_{it}}{\partial\theta_f\partial\theta_g^\top}\\ \left(\frac{\partial^2 m_{it}}{\partial\theta_f\partial\theta_g^\top}\right)^\top & \frac{\partial^2 m_{it}}{\partial\theta_g\partial\theta_g^\top}\end{pmatrix},$$

where

$$\begin{aligned}\frac{\partial^2 m_{it}}{\partial\theta_f\partial\theta_f^\top} &= -\frac{\partial^2 f(k_{it},v_{it};\theta_f)}{\partial\theta_f\partial\theta_f^\top}+\frac{\partial g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}}\frac{\partial^2 f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f\partial\theta_f^\top}\\ &\quad -\frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}^2}\left(\frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}\right)\left(\frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}\right)^\top,\\ \frac{\partial^2 m_{it}}{\partial\theta_f\partial\theta_g^\top} &= \frac{\partial f(k_{it-1},v_{it-1};\theta_f)}{\partial\theta_f}\frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\omega_{it-1}\partial\theta_g^\top},\\ \frac{\partial^2 m_{it}}{\partial\theta_g\partial\theta_g^\top} &= \frac{\partial^2 g(\mathrm{E}[q_{it-1}|x_{it-1}]-f(k_{it-1},v_{it-1};\theta_f);\theta_g)}{\partial\theta_g\partial\theta_g^\top},\end{aligned}$$

do not depend on the lagged prediction error ζ_{it-1}. Following standard asymptotic arguments, Γ can be consistently estimated by the finite-sample analog to equation (16) using the estimate of $\mathrm{E}[q_{it-1}|x_{it-1}]$ from step 1 and the estimate of θ from step 2.

References

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

Ackerberg, D. & De Loecker, J. (2024), Production function identification under imperfect competition, Working paper, University of Texas, Austin.

Andrews, I., Gentzkow, M. & Shapiro, J. (2017), 'Measuring the sensitivity of parameter estimates to estimation moments', *Quarterly Journal of Economics* **132**(4), 1553–1592.

Aw, B., Roberts, M. & Xu, D. (2011), 'R&D investment, exporting, and productivity dynamics', *American Economic Review* **101**(4), 1312–1344.

Becker, R., Gray, W. & Marvakov, J. (2021), 'NBER-CES manufacturing productivity database (1958-2018, version 2021a)', https://www.nber.org/research/data/nber-ces-manufacturing-industry-database.

Berry, S., Levinsohn, J. & Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**(4), 841–890.

Biondi, F. (2022), Firm productivity and derived factor demand under variable markups, Working paper, Katholieke Universiteit Leuven, Leuven.

Blundell, R. & Bond, S. (2000), 'GMM estimation with persistent panel data: An application to production functions', *Econometric Reviews* **19**, 321–340.

Brand, J. (2020), Estimating productivity and markups under imperfect competition, Working paper, University of Texas, Austin.

Cai, L., Guo, X. & Zhong, W. (2024), 'Test and measure for partial mean dependence based on machine learning methods', *Journal of the American Statistical Association* **forthcoming**.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018), 'Double/debiased machine learning for treatment and structural parameters', *Journal of Econometrics* **21**(1), C1–C68.

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

Chernozhukov, V., Hansen, C. & Spindler, M. (2015), 'Valid post-selection and post-regularization inference: An elementary, general approach', *Annual Review of Economics* **7**, 649–688.

Cobb, C. & Douglas, P. (1928), 'A theory of production', *American Economic Review* **18**(1), 139–165.

De Loecker, J. (2013), 'Detecting learning by exporting', *American Economic Journal: Microeconomics* **5**(3), 1–21.

De Loecker, J., Goldberg, P., Khandelwal, A. & Pavcnik, N. (2016), 'Prices, markups and trade reform', *Econometrica* **84**(2), 445–510.

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

Delgado, M. & Manteiga, W. (2001), 'Significance testing in nonparametric regression based on the bootstrap', *Annals of Statistics* **29**(5), 1469–1507.

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

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

Doraszelski, U. & Jaumandreu, J. (2024), Reexamining the De Loecker & Warzynski (2012) method for estimating markups, Working paper, University of Pennsylvania, Philadelphia.

Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', *American Economic Review* **98**(1), 394–425.

Gandhi, A., Navarro, S. & Rivers, D. (2020), 'On the identification of gross output production functions', *Journal of Political Economy* **128**(8), 2973–3016.

Hahn, J., Liao, Z. & Ridder, G. (2018), 'Nonparametric two-step sieve M estimation and inference', *Econometric Theory* **34**(6), 1281–1324.

Hall, R. (1988), 'The relation between price and marginal cost in U.S. industry', *Journal of Political Economy* **96**(5), 921–947.

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

Hu, Y. & Schennach, S. (2008), 'Instrumental variable treatment of nonclassical measurement error models', *Econometrica* **76**(1), 195–216.

Jaumandreu, J. (2022), The remarkable stability of us manufacturing markups, Working paper no. 17490, CEPR, London.

Jaumandreu, J. & Lopez, R. (2024), 'Markups in US food manufacturing accounting for non-neutral productivity', *Journal of Applied Economics* **75**(2), 573–591.

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

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

Mundlak, Y. & Hoch, I. (1965), 'Consequences of alternative specifications in estimation of Cobb-Douglas production functions', *Econometrica* **33**(4), 814–828.

Neyman, A. (1959), Optimal asymptotic tests of composite statistical hypotheses, *in* V. Grenander, ed., 'Probability and Statistics', John Wiley & Sons, New York, pp. 213–234.

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

Pakes, A. (1994), Dynamic structural models, problems, and prospects: mixed continuous discrete controls and market interactions, *in* C. Sims, ed., 'Advances in econometrics: Sixth World Congress', Vol. 2, Cambridge University Press, Cambridge.

Parente, P. & Santos Silva, J. (2016), 'Quantile regression with clustered data', *Journal of Econometric Methods* **5**(1), 1–15.

Ponder, M. (2021), Production functions without control functions: Estimating input elasticities when monotonicity does not hold, Working paper, University of Minnesota, Minneapolis.

Shah, R. & Peters, J. (2020), 'The hardness of conditional independence testing and the generalised covariance measure', *Annals of Statistics* **48**(3), 1514–1538.

Solow, R. (1957), 'Technical change and the aggregate production function', *Review of Economics and Statistics* **39**(3), 312–320.

Syverson, C. (2001), Market structure and productivity, Ph.d. dissertation, University of Maryland, College Park.

Van Biesebroeck, J. (2007), 'Robustness of productivity estimates', *Journal of Industrial Economics* **55**(3), 529–569.

von Thünen, J. (1842), *Der isolirte Staat in Beziehung auf Landwirthschaft und Nationalökonomie*, Leopold, Rostock.

Wicksteed, P. (1894), *An essay on the coordination of the laws of distribution*, Macmillan, London.