Production Function Estimation with Measurement Error in Capital

Charlotte Ursenbach
Western University

Job Market Paper

October 2026
The latest version can be found here

Abstract

Capital is notoriously hard to measure, and measurement error in capital can substantially bias estimates of production functions and productivity. Although this problem has been discussed for decades, no consensus exists on how best to correct for it. I develop a nonparametric firm-level production function estimator that corrects for capital measurement error by leveraging a "perpetual inventory" assumption on capital accumulation that is typical in the literature. Applying it to Chilean manufacturing data, I find that standard approaches systematically understate the output elasticity of capital and overstate how much productivity varies across firms.

1 Introduction

The estimation of production functions has become a major area of research interest with a wide variety of applications[1] with considerable research in particular on understanding the determinants of firm-level productivity.[2] Estimation of production functions generally relies on specialized estimators[3] to control for endogeneity between inputs and productivity. But like most estimators, production function estimators are vulnerable to issues of data quality. Measurement error in any input variable renders typical production function estimators inconsistent. Because inputs are typically correlated with each other, measurement error in any input has the potential to contaminate estimated output elasticities with respect to all inputs. Estimated productivity may also be affected, as variation in output that cannot be attributed to variation in inputs is attributed to productivity (Syverson, 2011).

Although there is widespread reliance within the production function literature on accurate input measurement, economists have long recognized in particular the difficulty of measuring capital stocks correctly (e.g., Bartelsman & Doms, 2000; Becker et al., 2006; Collard-Wexler & De Loecker, 2024; Hicks, 1981; Kim, Petrin, & Song, 2016; Mueller, 2008; Stevens, 1994). But while several approaches (see Section 2.2 for an overview) have been proposed for estimating production functions in the presence of measurement error in capital, none have seen widespread adoption. One particular hurdle is that the traditional instrumental variables (IV) method relies on linearity assumptions which are violated by the widely-used perpetual inventory method for measuring capital stocks.

In this paper, I develop and implement a production function estimator based on the nonlinear, non-classical measurement error correction of Hu and Schennach (2008, HS) that corrects for measurement error in capital when capital is measured by the perpetual inventory method. I assume the initial stock of capital suffers from measurement error; that error propagates to subsequent periods through a perpetual inventory assumption. The HS correction can be thought of as a nonparametric generalization of instrumental variables, requiring an instrument (or auxiliary measure) for capital; I use the firm's initial choice of intermediate inputs as an instrument for the firm's true initial capital. Intuitively, firms which use more material or energy inputs are likely to have more capital as well. I show that initial intermediates are a valid instrument for capital under a typical model of production augmented with mild additional assumptions on the distributions of measurement error and other inputs. Because HS-based estimators tend to have severe requirements in terms of data and computation, I provide a set of additional simplifying assumptions that render my estimator more tractable. Applying my estimator to Chilean manufacturing data, I find that, by failing to control for measurement error, standard methods can underestimate average capital elasticity by up to 49.9% and up to 31.2% of productivity variance obtained using standard

[1] Production function estimation has been used to estimate markups (e.g., De Ridder, Grassi, & Morzenti, 2026), the skill bias of technological change (e.g., Doraszelski & Jaumandreu, 2018), misallocation of inputs between firms (e.g., Sivadasan, 2009), and firm productivity heterogeneity.

[2] Research into productivity has considered factors such as research and development (e.g., Aw, Roberts, & Xu, 2008), institutions (e.g., Dollar, Hallward-Driemeier, & Mengistae, 2005), trade (e.g., Amiti & Konings, 2007; Bernard et al., 2003; Halpern, Koren, & Szeidl, 2015; Kasahara & Rodrigue, 2008), and trade barriers (e.g., De Loecker, 2011; Pavcnik, 2002; Topalova & Khandelwal, 2011).

[3] e.g., Ackerberg, Caves, and Frazer (2015), Gandhi, Navarro, and Rivers (2020), Levinsohn and Petrin (2003), Olley and Pakes (1996), Wooldridge (2009)

methods can be attributed to measurement error in initial capital.

A modification of the HS method was previously applied to the production function context by Kim, Petrin, and Song (2016), who use it to correct a control function estimator for capital measurement error without restricting the control function to be linear, by finding instruments for capital in each period. By focusing on initial measurement error, my estimator addresses production function estimation in the perpetual inventory context, where it is difficult to find valid instruments for capital in each period, but measurement error is propagated from one period to the next in a known way; this structure allows me to locate the measurement error in initial capital stocks, so that only a single instrument is necessary.

The structure my estimator relies on follows from a common approach to measuring capital. While it is often possible to obtain information from firms on the accounting or "book" value of their capital stock, this does not necessarily correspond to the economic concept of capital. For instance, book values of assets may reflect historical purchase prices rather than current valuations, and accounting depreciation rules applied by firms do not necessarily conform to the true loss of economic value. A common alternative approach is to construct estimates of capital recursively through what is known as the perpetual inventory method: in each period, capital is taken to be the previous period's capital minus depreciation, plus any investment. The perpetual inventory construction reflects a standard "law of motion" assumption on the accumulation of capital. To initialize the perpetual inventory procedure, an assumption must be made about the firm's initial capital stock; a typical choice in the literature is to use the firm-reported book value of capital for the first period and construct all subsequent periods with perpetual inventory. This approach reduces dependence on firm-reported capital stock in exchange for greater reliance on firm-reported investment flows, which are less vulnerable to common concerns about the accuracy of capital stocks. However, no matter how accurately the perpetual inventory method reflects the reality of capital accumulation, any error in a firm's assumed initial capital stock will propagate to all future periods.

The error propagated by the perpetual inventory method undergoes repeated nonlinear transformations that make it difficult to find instruments for mismeasured capital that are both relevant and excluded. I reduce the demands on a potential instrument by assuming that measurement error enters only through mismeasurement of the initial capital stock for each firm, before the perpetual inventory method introduces a relationship between error and other variables. This is equivalent to an assumption that the perpetual inventory equation accurately reflects the process of capital accumulation.

In order to handle this nonlinear error propagation and to admit nonlinear production functions, my estimator applies the HS measurement error correction technique, permitting me to estimate production functions nonparametrically. Intuitively, the HS approach works by jointly identifying three densities: one for the distribution of measurement error, one for the distribution of the instrument conditional on initial capital, and one for the distribution of output conditional on current and lagged inputs. The conditional density of output is written to depend on two structural functions of interest: the production function and a productivity process that controls the evolution of productivity over time. In an error-free setting, the first-order conditions for maximum likelihood estimation of the output density resemble the moment conditions used in the GNR production function estimator (Gandhi, Navarro, & Rivers, 2020).

Limiting the scope of my investigation to initial measurement error allows me to rely on first-period intermediates as an instrument for the entire time series of each firm, so that subsequent observations of intermediates can be used to control for endogeneity. I show that timing assumptions typical in the production function literature are suggestive of the validity of initial intermediates as an instrument for initial capital. Because I require fewer instruments, I can use each firm's entire time series to identify its single initial measurement error, rather than having to identify a separate measurement error in each period, easing my data and computational requirements. One implication of my assumption on measurement error is that investment is correctly measured. This assumption, shared by Collard-Wexler and De Loecker (2024), can be viewed as a judgment that investment, as a flow variable, may be measured more accurately than the stock of capital. Collard-Wexler and De Loecker leverage the assumption that investment is measured without error to control for a more general model of capital measurement error to a first-order approximation; by contrast, I control for one major source of capital measurement error exactly.

The resulting estimator is identified, but requires identifying a high-dimensional joint density function; in my implementation, I impose additional structure on the model to produce a simplified, tractable estimator. Most significantly, I impose Markovian assumptions on the distribution of investment and labor that allow me to decompose a high-dimensional joint density into a product of low-dimensional densities. This simplification allows me to limit the model to a manageable number of parameters.

Because most existing work in production function estimation uses the Generalized Method of Moments (GMM), whereas HS is a maximum likelihood estimation (MLE) method, I construct a variant of my estimator that imposes moment conditions similar to those used by GNR for greater comparability with the literature. The results from both methods are similar, so I present the moment-based estimates as my main results for greater interpretability.

I show through Monte Carlo simulations that, under a plausible data-generating process, initial measurement error can lead to severe estimation error, including substantial overestimates of the variance of productivity and large biases in the estimated production function, which my estimator can correct. I also demonstrate that my estimator can be robust to error that enters through channels other than the initial stock of capital.

Finally, I apply my estimator to Chilean manufacturing data, comparing it to GNR, a widely-used production function estimator which does not correct for measurement error in capital. I find that initial measurement error can result in severely biased estimates of productivity and production functions in the absence of error-correction, particularly with a nonlinear production function. Failing to correct for measurement error underestimates average capital elasticities by as much as 49.9%, and up to 31.2% of productivity variation estimated by the error-uncorrected method can be attributed to measurement error in capital. The estimated average labor elasticity rises to compensate for the downward bias in capital elasticity; the bias in labor elasticity is typically comparable in magnitude to the bias in capital elasticity, so that returns to scale are affected only mildly. While average intermediate elasticity is fairly robust to capital measurement error, variance in all three input elasticities (capital, labor, and intermediates) can be severely understated, by nearly a factor of 3.5 in the case of capital elasticity.

Because the determinants of firm productivity are of particular interest to researchers, I also compute correlations between productivity and four observable firm-level variables

excluded from the model: whether the firm exports, whether the firm imports, whether the firm advertises, and whether the firm pays above-median wages. My error-corrected estimator yields substantially different correlations from error-uncorrected GNR. In some cases, correlations estimated using the error-uncorrected estimator have opposite signs from the corresponding error-corrected correlations. Unlike the biases in average capital and labor elasticities, which tend to have predictable signs, the direction of the bias in productivity correlations is not generally consistent across industries, making it difficult to interpret error-uncorrected estimates.

The organization of the rest of the paper is as follows. In Section 2, I discuss measurement error in capital and review existing corrections. In Section 3, I describe a typical setting for production function estimation, discuss the source of identification for existing methods and motivate my approach by showing how the transmission of measurement error through perpetual inventory can cause both error-naive and linear IV error-correction methods to fail. I present my estimator and give details of its implementation in Section 4. I discuss practical issues involved in estimation in Section 5, and in Sections 6 and 7, I present my Monte Carlo and empirical results respectively. Finally, I conclude in Section 8.

2 Capital Measurement Error

In this section, I discuss why capital measurement error is a particular concern, and review the existing methods used to accommodate capital measurement error.

2.1 Measuring Capital

Most work on the estimation of production functions assumes that inputs are measured without error. However, the issue of measurement error in capital stock has been raised repeatedly for decades. "Capital stocks," write Bartelsman and Doms (2000), "are notoriously difficult to measure at a micro level." Becker et al. (2006) discuss substantial discrepancies between firm-level and sector-wide estimates of capital stocks. Within the production function literature, Griliches and Mairesse (1995) draw attention to the possibility of bias in production function estimates introduced by mismeasured capital; Levinsohn and Petrin (2003) note that "measurement error problems... tend to be most pronounced in capital"; and Mueller (2008) calls capital measurement "the major difficulty" in correctly measuring inputs. These difficulties may be highly consequential for the study of productivity as well as for production function estimation: as Syverson (2011) observes, "any output driven by unmeasured input variations... will show up as productivity."

One straightforward approach to measuring capital stock is to ask firms about the accounting or "book" value of their capital in each period. However, as the OECD manual on measuring capital (OECD, 2009) discusses, there are several known issues with the book value method: "data on company assets may be valued at historical costs" rather than at current prices; depreciation schedules used by firms may not reflect an actual "loss in market value"; and "fiscal considerations or fiscal rules such as accelerated depreciation" may spuriously impact capital valuations.

A common alternative method for estimating firm capital stocks is known as the perpetual

inventory method. Given capital (K) in the previous period $t-1$, investment (I) to be realized in the present period, and an assumed depreciation rate δ, the level of capital in the current period t can be constructed as

$$K_t = (1 - \delta)K_{t-1} + I_t \tag{1}$$

Therefore, given δ, an initial-period estimate K_0, and a stream of investments $I_1 \dots I_T$, the perpetual inventory method recursively constructs a time series of capital $K_1 \dots K_T$. As the period index t increases, the impact of K_0 on K_t shrinks.

One advantage of using Equation (1) to construct capital is that it mirrors a typical assumption about the law of motion for capital accumulation. To the extent that (1) accurately reflects the way capital accumulates, measurement error will only enter the capital series through the initial-period estimate, and will produce increasingly accurate estimates of capital as t increases. However, even under the assumption that the perpetual inventory method accurately reflects the law of motion for capital, it can only mitigate measurement error, since capital estimates are always contaminated by the initial estimate K_0.

Stevens (1994), commenting that "virtually all researchers acknowledge that the starting values [of capital] are subject to large errors," shows that for an inaccurate choice of K_0, regression estimates can be badly biased in the macroeconomic context. A typical approach in the microeconomic literature is to use the initial book value of capital as the starting estimate, meaning that the unreliability of book values still partially contaminates capital constructed through perpetual inventory. Even so, perpetual inventory and book value measures of capital can differ substantially (Becker et al., 2006; Collard-Wexler & De Loecker, 2024), which underscores the difficulty of constructing accurate capital estimates.[4]

Correcting for measurement error in capital is complicated by the distributional implications of the perpetual inventory approach. A traditional method of measurement error correction (first described in Wald, 1940) uses an instrumental variable regression involving an instrument for the mismeasured variable, which can alternatively be interpreted as an additional independent measurement. The IV error correction assumes that the mismeasured variable (in this case, capital) suffers from classical measurement error, meaning that it has an additively separable measurement error which is uncorrelated with both the true value and the instrument. One difficulty in adapting this method to the production function context is that, as I discuss in Section 3, even if measurement error is classical in the initially observed capital stock, the application of Equation (1) will generally introduce a relationship between measurement error and true capital.[5]

[4] One method for reducing measurement error used in macroeconomics is to "burn in" the perpetual inventory estimates by discarding the first decade (or more) of data, so that the starting estimate has only a small impact on estimated capital in the first retained period. This is not necessarily practical in the context of firm-level estimation. Production function estimation is usually conducted on relatively short panels, and within those panels firm exit and entry is common, so a lengthy burn-in period could eliminate a huge portion of the data and introduce selection bias by eliminating all short-lived firms, which may have a systematically different distribution of productivity, analogous to the issues with panel balancing discussed in Olley and Pakes (1996).

[5] As Collard-Wexler and De Loecker argue, this difficulty cannot necessarily be escaped by using firm-reported book values instead of perpetual inventory, since the firm may also be using perpetual inventory internally in computing their own measures of capital.

While econometricians have continually flagged concerns about mismeasurement of capital, no approach to measurement error correction has become widely accepted in the literature. I attempt to fill this gap by presenting an estimator that imposes the structural assumption that the perpetual inventory equation reflects the true process of capital accumulation in order to fully correct for measurement error in initial capital K_0. In this respect, it is a natural extension of a common choice to prefer perpetual inventory-based capital stocks over book values or other alternative measures of capital.

2.2 Existing Capital Measurement Error Corrections

Existing solutions to capital measurement error in the production function context can be broadly categorized by the assumptions they make on the distribution of measurement error. One family of methods assumes that capital suffers from classical measurement error. While measurement error is likely to be non-classical in general, as Collard-Wexler and De Loecker (2024) show, measurement error in capital, under a perpetual inventory framework, is classical under a log-linear approximation when investment is correctly measured. An advantage of approaches in this family is that they tend to have limited requirements in terms of data and computation.

Collard-Wexler and De Loecker (2024) and Galuscak and Lizal (2011) both adapt the traditional IV correction for measurement error (first described in Wald, 1940) to the production function setting. As mentioned above, these methods require the existence of an instrument correlated with true capital and uncorrelated with measurement error; Collard-Wexler and De Loecker use investment while Galuscak and Lizal use depreciation. Gorodnichenko (2007) proposes a novel structural estimator that compensates for classical measurement error in all inputs and prices (including capital) in the context of imperfectly competitive markets using several linear approximations.

In order to exploit the advantages of classicality in measurement error, methods in this family rely on strict assumptions about linearity. My estimator permits a nonlinear production function and productivity evolution, at the cost of added complexity. By departing from linearity and assuming that capital evolves according to perpetual inventory, I am able to handle the evolution of measurement error exactly, rather than relying on an approximation.

A second family of estimators relaxes the assumptions of classical measurement error. The estimator described in Kim, Petrin, and Song (2016) does not require noncorrelation between capital and measurement error. Like mine, their estimator builds on Hu and Schennach, but their work is different in several key respects. Their choice of instrument imposes an assumption that the current and lagged values of observed capital are independent of the second and third lags of observed capital after conditioning on true capital (and other observables), ruling out the possibility of serial correlation in measurement error. While alternative choices of instrument may resolve this problem, it is difficult to find a valid instrument in the perpetual inventory context, as I discuss in Section 3.4. However, their estimator can correct for within-firm measurement error, such as mismeasured investment, while mine assumes that the evolution of error over time is determined by the perpetual inventory equation.

Closely related to HS is Schennach (2004), which Dong and Hsieh (2021) use to consider capital measurement error. They treat energy use as an auxiliary measurement of capital; they assume energy intermediates are proportional to true capital plus an additive-in-levels

structural error. Because the auxiliary measure in Schennach (2004) plays a role similar to that of the instrument in HS, their energy auxiliary is analogous to my intermediate instrument.

Estimators in the HS family require nonparametric estimation of several density functions, which Collard-Wexler and De Loecker (2024) suggest may demand implausibly large datasets.[6] Because my estimator faces the same challenge, in implementing my estimator, I impose additional distributional assumptions in order to avoid having to nonparametrically estimate an intractably high-dimensional density. These distributional assumptions are not required for identification, so they can be relaxed in accordance with the available data. Additionally, by focusing its attention on the error that enters through initial capital, my estimator is able to use a firm's entire time series of observations to identify initial measurement error, rather than only using the current and previous periods of observations to identify current and previous measurement error. By limiting the scope of measurement error and imposing assumptions on the distribution of the data, I implement an estimator which is able to produce accurate results on typical-sized datasets, as I demonstrate via Monte Carlo.

In the next section, I show why perpetual inventory poses a particular challenge for corrections in the first family, which motivates the approach I take in this paper.

3 Setting and Motivation

In this section I describe a typical setting for production function estimation, based on GNR, and use it to develop intuition about the impact of measurement error. I first set out the model and show how GNR recover productivity and its innovation from inputs and output. I then add a perpetual inventory law of motion for capital and show that it causes initial measurement error to persist and become non-classical. Finally, I show how measurement error invalidates the GNR moment conditions, and why linear IV corrections are difficult to apply when capital is constructed by perpetual inventory.

3.1 Production and Productivity

In this subsection I describe a typical model of production based on GNR.

The econometrician observes firms $j = 1, \ldots, J$ over periods $t = 1, \ldots, T$. Firms use capital K^*_{jt}, labor L_{jt}, and intermediates M_{jt} to produce output Y_{jt}; log values of these quantities are denoted $k^*_{jt}, \ell_{jt}, m_{jt}, y_{jt}$, respectively. The log inputs are collectively referred to as $(k^*_{jt}, \ell_{jt}, m_{jt}) = x^*_{jt}$. Following the notation in HS, the asterisks on k^*_{jt} and x^*_{jt} indicate that these are true values. Later, I will use k_{jt} and x_{jt} to denote observed capital and $(k_{jt}, \ell_{jt}, m_{jt})$, respectively; this notation reflects an assumption that capital may be mismeasured, but other observed variables are not.

There is an asymptotically large population of firms from which to sample, so that $J \to \infty$ for fixed T; i.e., the panel can be short. Each firm's information set in period t is denoted

[6] Some authors have developed approaches that do not assume classical measurement error without building on HS. For instance, Kátay and Wolf (2008) assume that wages can serve as a perfect proxy for capacity utilization, under the assumption that the firm's wage premium reflects the cost of inducing longer hours or higher effort from employees in order to utilize capital more intensely, and Demirer (2023) achieves a set-identified production function estimator under very mild assumptions on the measurement error distribution.

$\mathcal{I}_{jt}$. The relationship of inputs to outputs is dictated by a production function F common to all firms, and a firm-specific productivity that can be broken into a persistent productivity ω_{jt}, and a transient productivity shock ε_{jt}, as described in (1):

Assumption 1. *The relationship between output and inputs takes the form*

$$Y_{jt} = F(K^*_{jt}, L_{jt}, M_{jt}) \exp(\omega_{jt} + \varepsilon_{jt})$$

or

$$y_{jt} = f(k^*_{jt}, \ell_{jt}, m_{jt}) + \omega_{jt} + \varepsilon_{jt} = f(x^*_{jt}) + \omega_{jt} + \varepsilon_{jt} \tag{2}$$

for some production function f which is differentiable, strictly increasing in m, and either linear or strictly concave in m, at all $x^* \in \mathbb{R}^3$.

ω_{jt} represents information known to the firm at the time it makes its period t decisions, while ε_{jt} represents unanticipated information. For instance, ω_{jt} may represent "the managerial ability of a firm, expected down-time due to machine breakdown, [or] expected defect rates in a manufacturing process" while ε_{jt} might represent "deviations from expected breakdown [or] defect" rates (Ackerberg, Caves, & Frazer, 2015). ε_{jt} may also represent measurement error in y_{jt}.

Assumption 2 formalizes this distinction and describes the statistical properties of ω_{jt} and ε_{jt}

Assumption 2. *$\omega_{jt} \in \mathcal{I}_{jt}$ is known to the firm at the time of making its period t decisions, while $\varepsilon_{jt} \notin \mathcal{I}_{jt}$ is not; both ε_{jt} and ω_{jt} are unobserved by the econometrician. ω_{jt} follows a first-order Markov process*

$$\omega_{jt} = h(\omega_{j,t-1}) + \eta_{jt} \tag{3}$$

with the productivity innovation η_{jt} independent of $\{\mathcal{I}_{j,t-1}, \varepsilon_{jt}\}$ and ε_{jt} independent of $\mathcal{I}_{jt}$. The function $h(\cdot)$ is continuous. Both ε_{jt} and η_{jt} admit densities denoted $f_\varepsilon(\cdot)$ and $f_\eta(\cdot)$, respectively.

Without loss of generality, normalize $\mathrm{E}[\varepsilon|\mathcal{I}_{jt}] = \mathrm{E}[\varepsilon] = 0$. For brevity I refer to h as the productivity process, although strictly speaking it only describes the evolution of persistent productivity. I depart slightly from GNR by assuming that the distributions of ε and η can be represented by density functions; this allows me to apply sieve maximum likelihood methods. The next assumption describes the market structure faced by firms.

Assumption 3. *Firms are price-takers in the output and intermediate-input markets, with $P^M_t \in \mathcal{I}_t$ denoting the period t price of intermediates and $P^Y_t \in \mathcal{I}_t$ the price of period t output. Firms maximize expected discounted profits.*

While Assumption 3 as written imposes that all firms share common prices, my identification still holds when there is observable firm-specific price variation, so long as the firms are price-takers.[7] For instance, observable variation in firm-specific prices due to heterogeneous

[7] In Online Appendix O6-4, GNR describe a variant of their estimator allowing for imperfect competition in output markets. While I do not consider this case in this paper, it is straightforward to modify my method in the same way.

transportation costs would provide additional identifying information that could be usefully incorporated into the estimator. However, a major advantage of GNR's approach is that it does not require firm-specific price variation, which is unavailable in many datasets. To retain this useful feature in my error-corrected estimator, I assume throughout that firms also face common prices.[8]

3.2 Recovering Productivity and Shocks

This section describes how GNR use Assumptions 1-3 to achieve identification, which is the foundation for the identification of my estimators. The approach of GNR identifies the production function f, the productivity process h, and (implicitly) the densities f_ε and f_η of the shocks ε and η, respectively. I collect these objects of interest into a parameter $\theta = (f, h, f_\varepsilon, f_\eta)$ taking values in a space Θ of functions and densities, with θ^0 denoting its true value.[9]

GNR's identification strategy uses the log of firm spending on intermediate inputs as a share of gross revenue, which they call s:

$$s_t(y, m) \equiv p_t^M - p_t^Y + m - y \tag{4}$$

Using first-order conditions on the firm's choice of intermediates, GNR show that log intermediate share can be written:

$$s_t(y_{jt}, m_{jt}) = \log\left(\frac{\partial}{\partial m} f(x_{jt}^*)\mathcal{E}\right) - \varepsilon_{jt} \tag{5}$$

where $\mathcal{E} = \mathrm{E}[\exp(\varepsilon)]$. Indexing s by t avoids carrying prices as separate arguments when prices are common to firms. Following GNR, I define

$$D^{\mathcal{E}}(x; \theta) \equiv \frac{\partial f}{\partial m}(x) \int \exp(\varepsilon) f_\varepsilon(\varepsilon)\, d\varepsilon \tag{6}$$

rewriting $\mathrm{E}[\exp(\varepsilon)]$ as a functional of f_ε. Solving Equation (5) for ε then gives

$$\tilde{\varepsilon}_t(y, x; \theta) \equiv \log D^{\mathcal{E}}(x; \theta) - s_t(y, m) \tag{7}$$

I use a tilde to distinguish this object, which is a function of θ, from the true shock ε_{jt}. It follows immediately that at the true value θ^0:

$$\tilde{\varepsilon}_t(y_{jt}, x_{jt}^*; \theta^0) = \varepsilon_{jt} \tag{8}$$

Because $\tilde{\varepsilon}_t$ depends on f only through $\partial f / \partial m$, Equation (5) cannot completely identify f by itself, but the partial derivative provides enough information to permit the innovation η to

[8] If firm-specific price information is available, my identification argument can easily be modified to allow for observed firm-specific prices.

[9] Note that θ is infinite-dimensional unless parametric restrictions are imposed on its components.

be written in terms of θ so as to pin down the remaining objects of interest. To do so, it is convenient to define

$$\mathcal{Y}_t(x;\theta) \equiv -\log D^{\mathcal{E}}(x;\theta) + p_t^M - p_t^Y + m - \int \frac{\partial f}{\partial m}(x)\, dm \tag{9}$$

$$\mathcal{C}(k,\ell;\theta) \equiv \int \frac{\partial f}{\partial m}(x)\, dm - f(x) \tag{10}$$

The integral $\int (\partial f/\partial m)\, dm$ is determined only up to a function of (k,ℓ); $\mathcal{Y}_t$ depends on f only through $\partial f/\partial m$, which is identified through (5), while $\mathcal{C}$ collects exactly the remaining information about f and is a function of (k,ℓ) alone. Substituting (4) and (7) into the sum of (9) and (10) gives

$$\begin{aligned}
\mathcal{Y}_t(x;\theta) + \mathcal{C}(k,\ell;\theta) &= -\log D^{\mathcal{E}}(x;\theta) + p_t^M - p_t^Y + m - f(x) \\
&= -\log D^{\mathcal{E}}(x;\theta) + y + s_t(y,m) - f(x) \\
&= y - \tilde{\varepsilon}_t(y,x;\theta) - f(x)
\end{aligned}$$

Persistent productivity can therefore be constructed as

$$\tilde{\omega}_t(x;\theta) \equiv \mathcal{Y}_t(x;\theta) + \mathcal{C}(k,\ell;\theta) \tag{11}$$

and by Assumption 1, $\tilde{\omega}_t(x_{jt}^*;\theta^0) = \omega_{jt}$. Finally, subtracting the conditional mean given the productivity process of Assumption 2 recovers the innovation:

$$\tilde{\eta}_t(x, x_{-1};\theta) \equiv \tilde{\omega}_t(x;\theta) - h\left(\tilde{\omega}_{t-1}(x_{-1};\theta)\right) \tag{12}$$

with $\tilde{\eta}_t(x_{jt}^*, x_{j,t-1}^*;\theta^0) = \eta_{jt}$.

The content of (7)–(12) is that a candidate θ delivers a complete set of reconstructed shocks for every observation, exact when $\theta = \theta^0$. GNR imposes moment conditions on the shocks $\tilde{\varepsilon}_t$ and $\tilde{\eta}_t$ in order to identify θ; I discuss those moments, and how measurement error in capital invalidates them, in Section 3.4. In Section 4.2, I use the method of HS to produce an estimator of θ which is robust to initial measurement error in capital. As an input to the HS method, I must first reimplement GNR's identification strategy as a maximum likelihood estimator in Section 4.1.

3.3 Perpetual Inventory

Assumption 1-3 are drawn almost directly from GNR; similar assumptions are common throughout in the production function literature. To consider measurement error in capital structurally, I add one more: that capital accumulates according to a perpetual inventory law of motion. I further assume that observed capital is constructed according to perpetual inventory, following the same law of motion as true capital:

Assumption 4. *Firms are endowed in period* 0 *with some stock of capital* K_{j0}^**. In period* t*, after observing* ω_{jt}*, firms purchase some level of investment* $I_{j,t+1}$ *which is realized in the subsequent period. Capital accumulates according to*

$$K_{j,t+1}^* = (1-\delta)K_{jt}^* + I_{j,t+1} \tag{13}$$

or equivalently,

$$K_{jt}^* = (1-\delta)^t K_{j0}^* + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau} \tag{14}$$

for some known $0 < \delta < 1$*. The econometrician knows the capital accumulation process, observes an initial* K_{j0}*, and constructs capital series according to*

$$K_{j,t+1} = (1-\delta)K_{jt} + I_{j,t+1} \tag{15}$$

or equivalently,

$$K_{jt} = (1-\delta)^t K_{j0} + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau} \tag{16}$$

It is common in empirical work for capital stocks to be constructed by perpetual inventory and assumed to be free of measurement error. Assumption 4 nests this setup by assuming $K_{j0} = K_{j0}^*$. However, when the econometrician observes only a noisy measure of initial capital, I will leverage Assumption 4 to identify each firm's entire series of capital from its initial observation. While Assumption 4, as written, represents the most common functional form for the perpetual inventory equation, it is sufficient that $K_{j,t+1}^*$ be a known function of K_{j0}^* and $I_{j1} \ldots, I_{j,t+1}$.

Assumption 4 has several important implications about measurement error. It is clear from Equations (14) and (16) that when the initial measure K_{j0} is measured with error, that error propagates to future periods of capital. In particular, if $\xi_{jt} = k_{jt} - k_{jt}^*$ is the measurement error in log capital, then

$$\xi_{jt} = \log \frac{(1-\delta)^t K_{j0} + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau}}{(1-\delta)^t K_{j0}^* + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau}} = \log \frac{(1-\delta)^t K_{j0}^* \exp(\xi_{j0}) + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau}}{(1-\delta)^t K_{j0}^* + \sum_{\tau=1}^{t}(1-\delta)^{t-\tau} I_{j\tau}} \tag{17}$$

When the depreciated sum of investments $\sum(1-\delta)^{t-\tau} I_{j\tau}$ is large relative to the depreciated initial capital $(1-\delta)^t K_{j0}$, ξ_{jt} is small, because the initial estimate of capital has little influence on the constructed K_{jt}. The use of perpetual inventory to construct capital stocks can therefore be motivated under Assumption 4 by the fact that after many periods of observed investment, the remaining measurement error in capital will be small.

Conversely, when the depreciated sum of investment at time t is small, then ξ_{jt} is close to ξ_{j0}. In particular, if the firm realizes zero investment in period t, then $\xi_{j,t-1} = \xi_{jt}$. In empirical datasets, it is common for investment to be zero in many periods, so the autocorrelation of ξ_{jt} over time is potentially very high.

ξ_{jt} is a nonlinear transformation of ξ_{j0} that depends on both true capital k_{j0}^* and I_{jt}. Outside of certain narrow special cases (such as when $I_{jt} = 0$ for all t), such a nonlinear transformation will introduce a relationship between ξ_{jt} and capital and investment even if one was not present in ξ_{j0}.[10] Because capital and investment are at least potentially corre-

[10]The objection could be raised that the nonlinearity of ξ_{jt} in terms of ξ_{j0} is an artifact of my choice to write the measurement error as additively separable in log capital. In particular, if error is treated as additively separable in levels, $K_{j0} = K_{j0}^* + \xi_{j0}$, then $K_{jt} = K_{jt}^* + \xi_{j0}$, and no correlation arises between measurement error and capital or investment. However, an assumption of additive separability in level of capital would have limited practical use. Exploiting classical measurement error assumptions on capital requires production functions which are linear in capital; the only production functions which are linear in level of capital are those for which capital is a perfect substitute for all other inputs.

lated with productivity, labor, and many other firm-level variables of interest, a correlation introduced between ξ_{jt} and capital or investment through Equation (16) can cause measurement error to become correlated with many observable quantities. As I discuss below, the artificially introduced relationship between ξ_{jt} and other observables poses a challenge for methods based on an assumption of classical measurement error, even when that assumption is valid in the initial observation k_{j0}.

3.4 The Impact of Measurement Error

Section 3.2 showed how a candidate θ reconstructs each firm's unobserved shocks from its inputs and output. Every one of those reconstructions is evaluated at true capital. Evaluating (11) at observed capital instead gives $\tilde{\omega}_t(x_{jt};\theta^0) \neq \omega_{jt}$ whenever $k_{jt} \neq k^*_{jt}$, and the reconstructed innovation $\tilde{\eta}_t$ inherits the discrepancy. In this subsection I discuss what that does to production function estimation. I focus on the GNR estimator, although analogous concerns apply to proxy-variable approaches. Other production function estimators, such as dynamic panel and index-number methods, may also be inconsistent in the presence of measurement error, but I do not consider them here.

The basic problem is straightforward. GNR identifies the production function and productivity process using moment conditions that require the productivity innovation to be uncorrelated with appropriately chosen functions of the firm's inputs and productivity history. These moment conditions are written in terms of the firm's true capital stock k^*_{jt}. When capital is measured with error, however, the econometrician must construct these moments using observed capital k_{jt}. The resulting measurement error can enter both the first and second stages of GNR and, in general, invalidates the moment conditions used for identification.

I first describe this problem in the linear Cobb-Douglas/AR(1) case, where the source of the problem has a particularly simple interpretation. In this special case, a subset of the second-stage moments of GNR can be written as the moment conditions of a two-stage least squares (2SLS) regression. This representation makes clear that measurement error creates an additional exclusion restriction: valid instruments must be uncorrelated not only with the productivity innovation, but also with the measurement error in capital. I then discuss why constructing such instruments is difficult when capital is measured using perpetual inventory. Finally, I return to nonlinear production functions and show that nonlinearities can introduce additional sources of inconsistency.

3.4.1 GNR moment conditions

The GNR estimator uses the following set of conditional moment conditions to achieve identification:

$$\mathrm{E}[\tilde{\varepsilon}_t(x^*_{jt}, y_{jt};\theta)|x^*_{jt}] = 0 \tag{18}$$

$$\mathrm{E}[\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1};\theta)|k^*_{jt}, \ell_{jt}, \mathcal{Y}_{j,t-1}(x^*;\theta), k^*_{j,t-1}, \ell_{j,t-1}] = 0. \tag{19}$$

Because $\tilde{\varepsilon}_t$ depends on partial information about f, specifically, on $\partial f/\partial m$, Equation (18) identifies $\partial f/\partial m$. Recall from (9) and (10) that the remaining information about f is col-

lected in $\mathcal{C}$, a function of (k, ℓ) alone, and that for any choice θ of the remaining parameters,

$$\tilde{\omega}_t(x^*_{jt}; \theta) = \mathcal{Y}_t(x^*_{jt}; \theta) + \mathcal{C}(k^*_{jt}, \ell_{jt}; \theta)$$

and

$$\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1}; \theta) = \tilde{\omega}_t(x^*_{jt}; \theta) - h(\tilde{\omega}_{t-1}(x^*_{j,t-1}; \theta))$$

reconstruct persistent productivity and its innovation. The conditional moments are typically converted into a just-identified system of unconditional moments:

$$\mathrm{E}\left[\tilde{\varepsilon}_t(x^*_{jt}, y_{jt}; \theta)\frac{(k^*_{jt})^{r_K}\ell^{r_L}_{jt}m^{r_M}_{jt}}{D(x^*_{jt})}\right] = 0 \tag{20}$$

$$\mathrm{E}\left[\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1}; \theta)(k^*_{jt})^{\tau_K}\ell^{\tau_L}_{jt}\right] = 0 \tag{21}$$

$$\mathrm{E}\left[\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1}; \theta)\mathcal{Y}^{\tau_Y}_{j,t-1}(x^*; \theta)\right] = 0. \tag{22}$$

Since moment condition (20) alone identifies a subset of the parameters, the usual procedure is to estimate the first-stage parameters using this condition and then estimate the remaining parameters using conditions (21)-(22). In the special case where the log production function is linear in intermediates, k^*_{jt} does not enter equation (20), so the first-stage parameters can be consistently estimated even when capital is measured with error. The more fundamental problem therefore arises in the second stage, where the measured capital stock enters the moments used to identify the production function and productivity process.

3.4.2 Measurement error in the linear case

To build intuition for this problem, consider the special case of a Cobb-Douglas production function and a linear AR(1) process for productivity,

$$f(k^*, \ell, m) = \alpha_K k^* + \alpha_L \ell + \alpha_M m$$

and

$$h(\omega) = \gamma_0 + \gamma_1 \omega.$$

Since the first-stage parameters can be estimated consistently in this linear case, suppose for now that they are known. For brevity, write $\mathcal{Y}_{jt} \equiv \mathcal{Y}_t(x^*_{jt}; \theta)$ for the realized value of (9) at firm j's period-t inputs. For any variable v_{jt} and any $\gamma \in \mathbb{R}$, define the quasi-differencing operator

$$\Delta_\gamma v_{jt} \equiv v_{jt} - \gamma v_{j,t-1}.$$

Intuitively, this operator transforms a variable by subtracting a lag weighted by the persistence parameter. At the true value γ_1, the GNR second-stage moments can then be written as

$$\mathrm{E}\left[\left(\Delta_{\gamma_1}\mathcal{Y}_{jt} - \alpha_K\Delta_{\gamma_1}k^*_{jt} - \alpha_L\Delta_{\gamma_1}\ell_{jt} - \gamma_0\right) \otimes q_{jt}\right] = 0 \tag{23}$$

$$\mathrm{E}\left[\left(\Delta_{\gamma_1}\mathcal{Y}_{jt} - \alpha_K\Delta_{\gamma_1}k^*_{jt} - \alpha_L\Delta_{\gamma_1}\ell_{jt} - \gamma_0\right) \mathcal{Y}_{j,t-1}\right] = 0 \tag{24}$$

where q_{jt} is a vector of instruments. The system (23)-(24) is converted into four unconditional moments used by GNR in the linear case by using $q_{jt} = (1, k^*_{jt}, \ell_{jt})$ as instruments in (23).

The first three of these moments have a familiar interpretation. Conditional on a choice of γ_1, Equation (23) for $q_{jt} = (1, k^*_{jt}, \ell_{jt})$ gives exactly the moment conditions defining a 2SLS regression of quasi-differenced output on quasi-differenced capital and labor, using k^*_{jt} and ℓ_{jt} as instruments. Thus, in the linear case, the GNR second stage can be viewed as a one-dimensional search over the productivity persistence parameter γ_1. For each candidate value γ, the production-function coefficients α_K, α_L can be obtained from the corresponding 2SLS regression, and γ_1 is chosen to satisfy Equation (24). More generally, for any generic instruments $q_{jt} = (1, q_{1,jt}, q_{2,jt})$, given a choice of γ_1, estimation of the production function coefficients can be written as a two-stage least squares regression, with the second-stage regression

$$\Delta_{\gamma_1} \mathcal{Y}_{jt} = \gamma_0 + \alpha_K \Delta_{\gamma_1} k^*_{jt} + \alpha_L \Delta_{\gamma_1} \ell_{jt} + \eta_{jt}$$

and corresponding first-stage equations

$$\Delta_{\gamma_1} k^*_{jt} = \beta^K_0 + \beta^K_K q_{1,jt} + \beta^K_L q_{2,jt} + u^K_{jt}$$

$$\Delta_{\gamma_1} \ell_{jt} = \beta^L_0 + \beta^L_K q_{1,jt} + \beta^L_L q_{2,jt} + u^L_{jt}.$$

where u^K_{jt}, u^L_{jt} are just the residuals from linear projection of the endogenous $\Delta_{\gamma_1} k^*_{jt}$, $\Delta_{\gamma_1} \ell_{jt}$ onto the vector of instruments, representing any variation in $\Delta_{\gamma_1} k^*_{jt}$, $\Delta_{\gamma_1} \ell_{jt}$ that is not (linearly) explained by q. As usual for two-stage least squares, consistent estimation requires that the instruments $(q_{1,jt}, q_{2,jt})$ be correlated with labor and capital but uncorrelated with the second-stage residual η_{jt}; under GNR this is satisfied by Assumption 2. This 2SLS interpretation is specific to the case where h is linear; it is useful here because it provides a simple way to understand how measurement error affects the GNR moments.

Now suppose that the econometrician observes, instead of k^*_{jt},

$$k_{jt} = k^*_{jt} + \xi_{jt},$$

where ξ_{jt} is measurement error. The second-stage regression at the true value of γ_1 is now

$$\Delta_{\gamma_1} \mathcal{Y}_{jt} = \gamma_0 + \alpha_K \Delta_{\gamma_1} k_{jt} + \alpha_L \Delta_{\gamma_1} \ell_{jt} + \left(\eta_{jt} - \alpha_K \Delta_{\gamma_1} \xi_{jt}\right).$$

and the corresponding first-stage equations are now

$$\Delta_{\gamma_1} k_{jt} = \beta^K_0 + \beta^K_K q_{1,jt} + \beta^K_L q_{2,jt} + \left(u^K_{jt} + \Delta_{\gamma_1} \xi_{jt}\right)$$

$$\Delta_{\gamma_1} \ell_{jt} = \beta^L_0 + \beta^L_K q_{1,jt} + \beta^L_L q_{2,jt} + u^L_{jt}.$$

This representation makes the source of the problem transparent. In the absence of measurement error, the instruments must be uncorrelated with the productivity innovation η_{jt}. With measurement error, however, the structural residual also contains

$$-\alpha_K \Delta_{\gamma_1} \xi_{jt}.$$

Consequently, valid estimation requires the additional exclusion restriction

$$\mathrm{Cov}(\Delta_{\gamma_1} \xi_{jt}, q_{1,jt}) = \mathrm{Cov}(\Delta_{\gamma_1} \xi_{jt}, q_{2,jt}) = 0.$$

These conditions are necessary for the instruments to be valid, in addition to the usual relevance conditions and the moment condition (24) required to identify γ_1.

For the usual GNR choice of instruments, these exclusion restrictions are not generally satisfied. If an instrument is a function of current or lagged observed capital, then it is difficult to rule out correlation between the instrument and the measurement error in capital. The problem is therefore not simply that measurement error makes capital an endogenous regressor. Rather, measurement error changes the error term of the quasi-differenced regression in a way that can invalidate the instruments used to identify the production-function coefficients.

This interpretation also clarifies why replacing the standard GNR instruments with alternative instruments is a natural potential solution. Suppose that one could find instruments q that were correlated with $\Delta_{\gamma_1} k_{jt}$ and $\Delta_{\gamma_1} \ell_{jt}$, while remaining uncorrelated with $\Delta_{\gamma_1} \xi_{jt}$. Such instruments would restore the exclusion restriction in the linear Cobb-Douglas/AR(1) case. This is analogous to the approaches taken by Collard-Wexler and De Loecker (2024) in the ACF context and Galuscak and Lizal (2011) in the LP context. In particular, Collard-Wexler and De Loecker show that under a first-order approximation, investment and labor can be valid choices for (q_1, q_2).

In the absence of such an approximation, finding an appropriate instrument is challenging. Even if initial measurement error ξ_0 is independent of the other sources of measurement error, subsequent measurement error is generated by the firm's history of capital and investment. In particular, second-period measurement error ξ_1 is a nonlinear transformation of initial measurement error ξ_0, initial capital k_0^*, and first-period investment I_0. Consequently, independence or noncorrelation between capital and measurement error in the initial period need not extend to subsequent periods. If measurement error is correlated with capital, it is also likely to be correlated with instruments that are relevant for capital. Thus, constructing instruments that are both relevant for the firm's capital stock and independent of its measurement error is particularly difficult when capital is constructed through perpetual inventory.

This difficulty motivates the approach taken in this paper. Rather than attempting to construct a linear IV estimator that satisfies these stringent measurement-error exclusion restrictions, I instead use an estimator based on HS. Before turning to that estimator, however, it is useful to consider why the linear case above understates the difficulties created by measurement error in a nonlinear production function.

3.4.3 Nonlinear production functions

The Cobb-Douglas/AR(1) case provides a useful benchmark because it isolates a relatively simple measurement-error problem. Nonlinearity can introduce additional sources of inconsistency. First, when the intermediate-input elasticity depends on capital, as in a translog or CES production function, the first-stage moment condition (20) itself depends on capital. Measurement error therefore contaminates the first-stage estimate of $\mathcal{Y}_{jt}$, so that the assumption that the first stage can be estimated correctly no longer holds even asymptotically.

Second, nonlinear terms in capital create additional measurement-error terms in the second stage. To see this, consider an especially favorable case in which measurement error is

perfectly persistent,

$$\xi_{jt} = \xi_{j,t-1},$$

and productivity is also perfectly persistent,

$$\omega_{jt} = \omega_{j,t-1}.$$

These assumptions are unlikely to hold in practice, but they will help to illustrate the distinction between the linear and nonlinear cases. Under these assumptions, consider first a linear production function in capital. The second stage can be represented as

$$\begin{aligned}\Delta\mathcal{Y}_{jt} &= \alpha_K \Delta(k^*_{jt} + \xi_{jt}) + \alpha_L \Delta\ell_{jt} + \eta_{jt} \\ &= \alpha_K \Delta k^*_{jt} + \alpha_L \Delta\ell_{jt} + \eta_{jt},\end{aligned}$$

where the persistent measurement error disappears under first differencing. The corresponding first-stage relationship can be written

$$\begin{aligned}\Delta k_{jt} &= \beta_0^K + \beta_K^K k_{jt} + \beta_L^K \ell_{jt} + u_{jt}^K \\ &= \beta_0^K + \beta_K^K k^*_{jt} + \beta_L^K \ell_{jt} + \left(u_{jt}^K + \beta_K^K \xi_{jt}\right).\end{aligned}$$

As long as the productivity innovation η_{jt} is uncorrelated with the measurement error, the measurement-error fixed effect is absorbed by the differencing operation, and this particular linear specification does not generate an additional source of bias.

This result does not extend to nonlinear terms in capital. Suppose, for example, that $\Delta(k_{jt}^2)$ is also an endogenous regressor. The second stage then contains

$$\Delta\mathcal{Y}_{jt} = \alpha_K \Delta(k^*_{jt} + \xi_{jt}) + \alpha_{KK} \Delta(k^*_{jt} + \xi_{jt})^2 + \alpha_L \Delta\ell_{jt} + \eta_{jt}.$$

Under the assumption that $\xi_{jt} = \xi_{j,t-1}$, this becomes

$$\Delta\mathcal{Y}_{jt} = \alpha_K \Delta k^*_{jt} + \alpha_{KK} (\Delta k^*_{jt})^2 + \alpha_L \Delta\ell_{jt} + \left(\eta_{jt} + 2\alpha_{KK}\xi_{jt}\Delta k^*_{jt}\right).$$

The measurement error therefore no longer disappears. The nonlinear transformation of capital creates an additional interaction between the measurement error and true capital which is not eliminated by the first-differencing. An instrument such as measured capital $k_{jt} = k^*_{jt} + \xi_{jt}$ will generally be correlated with this additional error term, so the IV exclusion restriction fails.

Thus, the linear Cobb-Douglas/AR(1) case provides a useful intuition for the general problem but is also unusually benign. With nonlinear production functions, measurement error can contaminate the first-stage estimate of output and generate additional errors through nonlinear functions of capital. Nonlinearity in the productivity process can provide another potential source of bias, although in my applications the degree of the productivity process has generally made little difference in practice.

The following section develops an estimator based on HS that avoids the need to construct instruments satisfying these stringent measurement-error exclusion restrictions.

4 Estimators

While the production function estimation literature usually relies on GMM estimators, the estimator of Hu and Schennach (2008) is designed to correct a maximum-likelihood estimator. I therefore begin by describing a maximum-likelihood estimator analogous to the GNR estimator, which I refer to simply as MLE. Next, I extend that estimator to handle measurement error in capital using the approach of Hu and Schennach; I refer to the resulting error-corrected estimator as MLE-HS. For greater comparability with the GMM estimators widely used in the literature, I also describe a two-step variant of MLE-HS that I call GMM-HS, which I use to generate my main estimates. For clarity, I focus on the cases where there are only as many periods as needed for identification: two in the case of MLE, and three in the case of MLE-HS and GMM-HS.

4.1 MLE

In order to exploit the HS error correction, I describe a maximum likelihood estimator inspired by GNR, which I refer to as MLE. I show that the joint density of y_t, m_t conditional on k_t^*, ℓ_t and x_{t-1}^* can be expressed as the product of densities of ε_t and η_t. This enables the nonparametric recovery of both the functions f and h and the densities f_ε and f_η of ε and η, respectively. Because the densities can be estimated sieve nonparametrically, it is not necessary to make any parametric assumptions about the forms of f, h, f_ε or f_η beyond mild regularity conditions. I conclude by discussing the close relationship between GNR and MLE.

It is convenient to define

$$z_t^* = \{k_{jt}^*, \ell_{jt}, x_{j,t-1}^*\} = \{k_{jt}^*, \ell_{jt}, k_{j,t-1}^*, \ell_{j,t-1}, m_{j,t-1}\}$$

The density of (y_{jt}, m_{jt}) conditional on z_{jt}^* can be factored by Bayes's rule into the product of two conditional densities:

$$f_{y,m|z^*}(y_{jt}, m_{jt}|z_{jt}^*) = f_{y|m,z^*}(y_{jt}|m_{jt}, z_{jt}^*) f_{m|z^*}(m_{jt}|z_{jt}^*) \tag{25}$$

The densities $f_{y|m,z^*}$ and $f_{m|z^*}$ correspond respectively to the two stages of GNR estimation. Just as the two stages of GNR recover ε_{jt} and η_{jt}, respectively, the two densities on the right-hand side of (25) can be written in terms of the densities of ε_{jt} and η_{jt}. The density $f_{y|m,z^*}(y_{jt}|m_{jt}, z_{jt}^*)$ conditions on all inputs as well as sufficient information to recover ω_{jt}, so by (2), all remaining variation in y_{jt} comes from ε_{jt}. The density $f_{m|z^*}$ conditions on capital and labor, and sufficient information to recover lagged productivity $\omega_{j,t-1}$, so that by the existence of an intermediate demand function $\mathbb{M}$, all remaining variation in m_{jt} comes from η_{jt}. The expressions of $f_{y|m,z^*}$ and $f_{m|z^*}$ in terms of the densities of ε_{jt} and η_{jt} will also depend on the unknown functions f and h, allowing joint maximum likelihood estimation of the four functions of interest, $\theta = (f, h, f_\varepsilon, f_\eta)$.

Recall from Section 3.2 that at the true value θ^0, the shocks ε and η can be recovered exactly using the functions $\tilde{\varepsilon}_t$ and $\tilde{\eta}_t$, which depend only on parameters θ and observations of output and inputs $y_{jt}, x_{jt}^*, x_{j,t-1}^*$.

Lemma 1. *Under Assumptions 1-3, the conditional density of y_{jt} given x^*_{jt},$x^*_{j,t-1}$ is*

$$f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt};\theta^0) = f_{\varepsilon,t}(y_{jt}, x^*_{jt};\theta^0) f_{\eta,t}\left(x^*_{jt}, x^*_{j,t-1};\theta^0\right) \left|\frac{\partial}{\partial m_{jt}}\mathcal{Y}_t(x^*_{jt};\theta^0)\right| \tag{26}$$

where

$$f_{\varepsilon,t}(y,x;\theta) \equiv f_\varepsilon(\tilde{\varepsilon}_t(y,x;\theta))$$
$$f_{\eta,t}\left(x,x_{-1};\theta\right) \equiv f_\eta\left(\tilde{\eta}_t\left(x,x_{-1};\theta\right)\right)$$

The proof of Lemma 1 is given in Appendix B. Given the explicit expression in Equation (39) for the density of y, m given z^*, there is a natural sieve-nonparametric estimator

$$\hat{\theta}_n = \underset{\theta\in\mathcal{A}_n}{\operatorname{argmax}} \frac{1}{NT}\sum_{j=1}^{N}\sum_{t=1}^{T} \log f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt};\theta) \tag{27}$$

The parameter space $\mathcal{A}_n$ of θ is discussed in Appendix B.2. Establishing the consistency of $\hat{\theta}_n$ as an estimator of θ^0 requires several mild regularity conditions. In Appendix B.2 I give sufficient conditions for the consistency of θ_n when $T = 2$.

The MLE estimator is closely related to the GMM estimator of GNR, because the likelihood is maximized by setting the gradient of the log-likelihood to zero (i.e., the score function), which generates a system of equations similar to the moments used by GNR. In particular, the second stage of GNR typically uses moments of the form $\mathrm{E}[\hat{\eta}_{jt}k_{jt}^{\tau_k}\ell_{jt}^{\tau_\ell}] = \mathrm{E}[\hat{\eta}_{jt}\hat{\mathcal{Y}}^a_{jt}] = 0$ or $\mathrm{E}[\hat{\eta}_{jt}k_{jt}^{\tau_k}\ell_{jt}^{\tau_\ell}] = \mathrm{E}[\hat{\eta}_{jt}\hat{\omega}^a_{jt}] = 0$ for different powers of τ_k, τ_ℓ, a. The corresponding first-order conditions on MLE, on the other hand, are:

$$\mathrm{E}\left[\frac{f'_\eta(\eta)}{f_\eta(\eta)}(k_{jt}^{\tau_k}\ell_{jt}^{\tau_\ell} - h'(\hat{\omega}_{j,t-1})k_{j,t-1}^{\tau_k}\ell_{j,t-1}^{\tau_\ell})\right] = 0$$

$$\mathrm{E}\left[\frac{f'_\eta(\eta)}{f_\eta(\eta)}\hat{\omega}^a_{j,t-1}\right] = 0$$

for powers of τ_k, τ_ℓ, a. Under normality, $f'_\eta(\eta)/f_\eta(\eta)$ reduces to a constant times η, and the MLE score function can therefore be interpreted as an alternative choice of unconditional moments satisfying the conditional moment condition $\mathrm{E}[\eta_{jt}|\mathcal{I}_{jt}] = 0$ that GNR uses to identify its second-stage parameters. In particular, in the case where f_η is normal and h is linear, the MLE score function, conditional on a choice of γ_1, is identical to the moments of the OLS estimator of $\Delta_{\gamma_1}\mathcal{Y}_{jt}$ regressed on $\Delta_{\gamma_1}k_{jt}$, $\Delta_{\gamma_1}\ell_{jt}$.[11]

[11] More generally, for any f_η, the gradient with respect to $\gamma_a, a > 0$ is $\frac{f'_\eta(\eta)}{f_\eta(\eta)}\hat{\omega}^a_{t-1}$, which is never strictly proportional to $\eta_{jt}\mathcal{Y}^a_{jt}$ and is strictly proportional to $\eta_{jt}\omega^a_{jt}$ only when f_η is normal. Consequently, it is impossible to choose a family of distributions for η such that the score of the maximum likelihood estimator coincides exactly with the typical moments used in the method of moments in finite samples.

4.2 MLE-HS

The MLE estimator of the previous subsection assumes that true capital is observed. I now describe a nonparametric estimator based on HS which accounts for measurement error in the initial period. I begin by defining additional notation and giving assumptions, adapted from HS, under which it is possible to nonparametrically identify the joint distribution of true initial capital and observed variables. This joint distribution is naturally partitioned into three conditional distributions, one of which depends explicitly on the production function and productivity process. Therefore, identifying the joint density of observables and true initial capital also identifies the production function and productivity process. While the relatively mild assumptions given in this section are sufficient to identify the densities of interest, in practice it is useful to impose additional structure on the estimator; I briefly discuss these issues here and return to them in detail in Section 5.1.

Define a set of dependent variables $\mathbf{u}_j = (y_{jt}, m_{jt}, I_{jt}, \ell_{jt})_{t=2}^{T-1}$ and conditioning variables $\mathbf{w}_j = (y_{j0}, y_{j1}, I_{j1}, \ell_{j0}, \ell_{j1}, m_{j1})$. I treat labor and investment as dependent variables, which allows them to provide further information to assist in the identification of initial capital and simplifies the exposition somewhat.

Assumption 5. *The joint density of $\mathbf{u}_j, k_0, k_0^*, m_0$ conditional on $\mathbf{w}$ exists and is bounded, and so are all the marginal and conditional densities.*

Assumption 6. *Let*

$$L_{b|k_0^*}[g](b, \mathbf{w}) = \int_{\text{supp}(k_0^*)} f_{b|k_0^*}(b|k_0^*, \mathbf{w}) g(k_0^*|\mathbf{w}) dk_0^*$$

for $b = k_0, m_0$ where g is any bounded absolutely integrable function with domain equal to $\operatorname{supp} k_0^*$. *The operators $L_{k_0|k_0^*}$ and $L_{m_0|k_0^*}$ are one-to-one for any $\mathbf{w}$*

Assumption 7. *For all $k_0^* \neq \tilde{k}_0^*$ and all $\mathbf{w}$, the set $\{\mathbf{u}_j : f_{u|k_0^*,w}(\mathbf{u}_j|k_0^*, \mathbf{w}) \neq f_{u|k_0^*,w}(\mathbf{u}_j|\tilde{k}_0^*, \mathbf{w})\}$ has positive probability.*

Assumption 8. *Measurement error in capital is mean zero and independent of all other variables.*

(i) $f_{u|k,k^*}(\mathbf{u}_j|k_{j0}, k_{j0}^*, \mathbf{w_j}) = f_{u|k^*}(\mathbf{u}_j|k_{j0}^*, \mathbf{w_j})$.

(ii) $f_{k|k^*,m_0}(k_{j0}|k_{j0}^*, \mathbf{w_j}, m_{j0}) = f_\xi(\xi_{j0})$

(iii) $\mathrm{E}[\xi_0] = 0$

Assumption 9. *$\mathbf{u}_j$ is independent of m_0 conditional on k_{j0}, k_{j0}^* and $\mathbf{w}_j$.*

Assumptions 5 and 6 are fairly weak regularity conditions; see Hu and Schennach (2008) and W. K. Newey (2013) for a discussion of the latter, which is a nonparametric rank condition. Assumption 7 requires that true capital matters "enough" to the endogenous variables $\mathbf{u}$; specifically, it requires that no two values of true initial capital k_0^* result in identical distributions $f_{y,m|z^*}$ conditional on $\mathbf{w}$.

Assumptions 8 and 9 imply HS Assumptions 2 and 5. Assumption 8(i) says that the measurement error in capital is unrelated to dependent variables $\mathbf{u}_j$ conditional on $\mathbf{w}_j$ and k^*_{j0}. Assumption 8(ii) says that measurement error in capital is unrelated to initial intermediates m_{j0}, true capital k^*_{j0} and $\mathbf{w}_j$. Assumption 8(iii) requires the measurement error to be mean zero.[12]

Assumption 9 says that investment, labor, intermediates, and output from period two onward do not depend on initial intermediates except through true capital and conditioning variables $\mathbf{w}_j$. Due to the firm's static first-order condition on intermediates, initial intermediates contain information on current-period prices, productivity, labor, and capital. Given initial labor (contained in $\mathbf{w}_j$) and initial true capital, and treating prices (which are assumed to be common across firms) as constants, initial intermediates only contain information on initial productivity. Moreover, given true capital, productivity in period one is a function of $\mathbf{w}_j$ and initial true capital through the invertibility of intermediate demand, and due to Assumption 2, conditioning on period one productivity eliminates the dependence of productivity in all future periods on initial productivity.

However, Assumption 9 is not automatically satisfied; because almost no assumptions have been made on the data-generating process for labor and investment, it is conceivable that initial productivity could be informative as to future values of $\mathbf{u}_j$ even after conditioning on $\mathbf{w}_j$ and true capital. For instance, if labor or investment were chosen two periods in advance instead of one, labor in period two would be informative about productivity in period zero (and hence, intermediates in period zero) even after conditioning on period one inputs, violating Assumption 9. That said, in any case where only a finite number of lags of productivity are relevant to the distribution of labor and investment, Assumption 9 can be satisfied by conditioning on additional periods of observables.

Consider the following expression for the joint density of dependent variables, true capital, and observed capital, conditional on initial intermediates and conditioning variables $\mathbf{w}_j$:

$$f_{u,k_0,k^*_0|m_0}(\mathbf{u}_j, k_{j0}, k^*_{j0}|\mathbf{w}_j, m_{j0}) = f_{u|k^*_0}(\mathbf{u}_j|\mathbf{w}_j, k^*_{j0}) f_\xi(\xi_0|k^*_0) f_{k^*_0|m_0}(k^*_{j0}|m_{j0}, \mathbf{w}_j) \tag{28}$$

It is easy to show that Equation (28) is true by factoring the density on the left-hand side into a product of conditional densities $f_{u|k^*_0,k_0,m_0} f_{k_0|k^*_0,m_0} f_{k^*_0|m_0}$ and applying Assumptions 8 and 9. Integrating over k^*_{j0} yields

$$\begin{aligned} f_{u,k_0|m_0}(\mathbf{u}_j, k_{j0}|\mathbf{w}_j, m_{j0}) &= \int_{\text{supp}(k^*_{j0})} f_{u,k_0,k^*_0|m_0}(\mathbf{u}_j, k_{j0}, k^*_{j0}|\mathbf{w}_j, m_{j0}) dk^*_0 \\ &= \int_{\text{supp}(k^*_{j0})} f_{u|k^*_0}(\mathbf{u}_j|\mathbf{w}_j, k^*_{j0}) f_\xi(\xi_0|k^*_0) f_{k^*_0|m_0}(k^*_{j0}|m_{j0}, \mathbf{w}_j) dk^*_0 \end{aligned} \tag{29}$$

Theorem 1. *Under Assumptions 1-9, Equation 29 uniquely determines the three densities on its right-hand side*

Proof. See Appendix B.3 □

[12] Assumption 8(ii)-(iii) could be relaxed considerably. For instance, the variance of ξ_{j0} could depend on the true value of capital or the observables $\mathbf{w}_j$ and m_{j0}, so long as the conditional mean of ξ_0 is zero. For a fuller treatment of the possibilities, see Hu and Schennach (2008).

As I show in Theorem 1, under Assumptions 1-9, Equation (29) uniquely determines the three densities on its right-hand side, so all three are identified. In particular, $f_{u|k_0^*}$ identifies $f_{y,m|z^*}$ because for $t > 1$, $y_{jt}, m_{jt} \in \mathbf{u}_j$ and z_{jt}^* is a known function of $\mathbf{u}_j, k_{j0}^*$, so $f_{y,m|z^*}$ is identified. By Lemma 1, $f_{u|k_0^*}$ therefore identifies θ. For instance, in the case where $T = 3$:

$$f_{u,k_0|m_0}(\mathbf{u}_j, m_{j0}, k_{j0}|\mathbf{w}_j; \theta) = \int_{\text{supp}(k_{j0}^*)} f_{y,m|z^*}(y_{j2}, m_{j2}|z_{j2}^*; \theta) f_{I,L}(I_{j2}, \ell_{j2}|k_{j0}^*, \mathbf{w}_j) \times f_\xi(\xi_0|k_0^*) f_{k_0^*|m_0}(k_{j0}^*|m_{j0}, \mathbf{w}_j) dk_0^* \tag{30}$$

While Equation 29 is identified, $f_{u|k_0^*}$ is potentially very high-dimensional, and high-dimensional densities can be challenging to estimate even in large finite samples. In particular, while my identifying assumptions mean that y_{jt}, m_{jt} depend only on a small subset of $\mathbf{u}_j, \mathbf{w}_j$, there are few restrictions on the joint distribution of $\{I_{jt}, \ell_{jt}\}_{t=2}^{T-1}|\mathbf{w}_j$. I address this dimensionality problem in Section 5.1, which imposes a set of simplifying assumptions on the data-generating process to factor the density $f_{u|k_0,k_0^*,m_0}$ into a product of relatively simple densities.

4.3 GMM-HS

As discussed above, the MLE estimator does not exactly align with the GMM estimator of GNR as it is typically used. Under normality of η, MLE can be interpreted as imposing the conditional moment condition of GNR using a different set of instruments. This poses two issues. First, recent literature (e.g., Ackerberg, 2023) has emphasized the sensitivity of estimates to the choice of unconditional moment, so implicitly imposing a different unconditional moment may hamper comparability. Second, when γ_1 is large, the variance of $\Delta_{\gamma_1} k_{jt}^*$ is likely to be small, and $\mathcal{C}$ may be weakly identified. In this section I show how the MLE-HS estimator can be modified to more closely track the moments used in GNR.

In the absence of error-free data, the sample moments used in GNR are not feasible to compute directly, but given an estimated distribution of error-free data, an analogous feasible moment condition can be established. It is useful to partition the parameters to be estimated into two parts. Let θ_1 be the vector of parameters necessary for error-free estimation – $f, h, f_\varepsilon, f_\eta$ – and let θ_2 be all remaining parameters estimated by MLE-HS. Then

$$\mathrm{E}\left[\int_{\text{supp}(k_0^*)} g(\mathbf{u}_j, \mathbf{w}_j, k_0^*; \theta_1)\, f_{k_0^*|u,k_0,m_0}(k_0^*|\mathbf{u}_j, k_{j0}, \mathbf{w}_j, m_{j0}; \hat{\theta}_1, \hat{\theta}_2)\, dk_0^*\right] = 0 \tag{31}$$

where

$$f_{k_0^*|u,k_0,m_0}(k_0^*|\mathbf{u}_j, k_{j0}, \mathbf{w}_j, m_{j0}; \hat{\theta}_1, \hat{\theta}_2) = \frac{f_{u,k_0,k_0^*|m_0}(\mathbf{u}_j, k_{j0}, k_0^*|\mathbf{w}_j, m_{j0}; \hat{\theta}_1, \hat{\theta}_2)}{f_{u,k_0|m_0}(\mathbf{u}_j, k_{j0}|\mathbf{w}_j, m_{j0}; \hat{\theta}_1, \hat{\theta}_2)} \tag{32}$$

is the density of true initial capital implied by the MLE-HS estimates, with numerator and denominator given by Equations (28) and (29), respectively, and $g(\cdot; \theta_1)$ represents the moment in question, evaluated at the capital series implied by k_0^* and observed investment through Assumption 4. Because MLE-HS is consistent, it can be used to generate a consistent estimate of θ_2, which is then held fixed as Equation (31) is used with moments from

Equations (20)-(22) and moments describing the distributions of f_ε, f_η to estimate θ_1. I then re-estimate θ_2 by maximum likelihood, holding θ_1 fixed.[13] This procedure is analogous to two-step GMM or the pseudo-maximum likelihood approach of Hotz and Miller (1993), discussed in Aguirregabiria and Mira (2007). It also shares similarities with the two-step approach of Kim, Petrin, and Song (2016). In Appendix C I give results estimated using maximum likelihood.

5 Estimation

The MLE and MLE-HS estimators described in the previous section are identified under relatively mild assumptions, but implementing them requires estimating high-dimensional joint densities. In this section I describe the additional structure I impose to make estimation tractable, and the sieve spaces I use to approximate the remaining unknown functions.

5.1 Distributional Assumptions

My error-correction approach requires the estimation of the densities of inputs conditional on the firm's entire history of inputs, which may involve a considerable number of variables. The $f_{k_0^*|m_0}$ term conditions initial capital on initial intermediates and the six conditioning variables in $\mathbf{w}_j$, while the $f_{u|k_0^*}$ term is a joint density of $4(T-2)$ variables conditional on initial capital and, again, the six conditioning variables in $\mathbf{w}_j$. In general, nonparametric estimation of high-dimensional densities can be demanding in terms of computation and data, but in practice, the dependence of current variables on deeply lagged variables may be limited. In this section I give an example of a set of distributional assumptions on inputs that can be used to limit the dimensionality of my estimator. These simplifying assumptions, which I collect in Assumption 10, below, are primarily of importance because they will allow me to factor the joint density of each firm's series of inputs into a product of simpler conditional densities, limiting the number of parameters to be estimated. They are not used in my proof of identification but rather represent a compromise between tractability and generality.

Concretely, Assumption 10 along with Bayes' rule delivers the factorization

$$f_{u|k_0^*}(\mathbf{u}_j|\mathbf{w}_j, k_{j0}^*) = \prod_{t=2}^{T-1} f_{I,L}(I_{jt}, \ell_{jt}|x_{j,t-1}^*) f_{y,m|z^*}(y_{jt}, m_{jt}|z_{jt}^*) \tag{33}$$

which replaces a $4(T-2)$-dimensional density with a product of densities in four variables apiece, each built from objects already defined in Section 4. A parallel factorization applies to $f_{k_0^*|m_0}$; I give both, along with the explicit form of $f_{I,L}$, after stating and discussing the assumption.

Assumption 10. *For $j = 1 \dots N$, and $t = 1 \dots T-1$, let $\zeta_{jt}^I, \zeta_{jt}^L, \zeta_j^K$ be continuous random variables i.i.d. over t and j with mean zero and positive finite variance. Assume $\zeta_{jt}^I, \zeta_{jt}^L \perp \mathcal{I}_{j,t-1}$ and $\zeta_{jt}^I \perp \zeta_{jt}^L$ for all j, t. In period t, firms choose the levels of labor and investment to be realized in period $t+1$ as follows*

[13] In principle I could continue iterating until convergence, but it is computationally demanding to do so. In practice, quantities of interest change very little when θ_2 is re-estimated.

(i) In period t, firms choose to invest in period $t+1$ with probability

$$\Pr(I_{j,t+1} > 0 | \mathcal{I}_{jt}) = \mathbb{G}(k^*_{jt}, \ell_{jt}, m_{jt}, t)$$

for some $\mathbb{G} : \mathbb{R}^4 \to [0,1]$

(ii) Firms that choose to invest set

$$I_{j,t+1} = \exp(\mathbb{I}(k^*_{jt}, \ell_{jt}, m_{jt}, t) + \zeta^I_{j,t+1})$$

for some $\mathbb{I} : \mathbb{R}^4 \to \mathbb{R}$. Otherwise, $I_{j,t+1} = 0$.

(iii) All firms choose next-period labor according to

$$\ell_{j,t+1} = \mathbb{L}(k^*_{j,t+1}, k^*_{jt}, \ell_{jt}, m_{jt}, t) + \zeta^L_{j,t+1}$$

for some $\mathbb{L} : \mathbb{R}^5 \to \mathbb{R}$.

(iv) All firms have initial endowments of true capital given by

$$k^*_{j0} = \mathbb{K}(\ell_{j0}, m_{j0}, t_{j0}) + \zeta^K_j$$

for some $\mathbb{K} : \mathbb{R}^3 \to \mathbb{R}$, where t_{j0} is the period in which firm j is first observed.

(v) $\zeta^I_{jt}, \zeta^L_{jt}, \zeta^K_j$ are normally distributed

The five parts do different kinds of work. Assumptions 10(i) and 10(ii) let investment be factored out of the joint density one period at a time, accommodating a possible mass point at zero investment; Assumption 10(iii) does the same for labor; Assumption 10(iv) supplies the distribution of the $f_{k^*|\ell,m}$ term; and Assumption 10(v) simplifies the parameter space by imposing distributional assumptions on the residuals $\zeta^I, \zeta^K, \zeta^L$.

Assumptions 10(i)-(iii) are loosely inspired by timing assumptions discussed in Ackerberg, Caves, and Frazer (2015). Firms begin with an initial draw of capital k^*_{j0}, which may be correlated with labor and intermediates. In each period, firms first choose whether to invest; a function $\mathbb{G}$ of time and current-period inputs determines the firm's investment probability. Those firms which choose to invest plan a level of investment $\mathbb{I}$, which is likewise a function of time and current inputs. After firms plan investment they realize a log-additively separable shock ζ^I_{jt} to their planned investment. This might reflect, for instance, optimization error or firm-specific transient frictions. In panels where every firm is observed to make a positive investment each year, it is trivial to replace Assumption 10(i) with an assumption that all firms invest with probability 1. This is the case in my Monte Carlo estimates, but not in my empirical data.

After observing the investment shock, firms plan a level of labor $\mathbb{L}$ to be realized in the next period. When planning labor, I assume the firm has already observed the investment shock ζ^I_{jt}, and therefore allow planned labor to be a function of both current-period inputs and next-period capital. After making the plan, firms then experience another unobserved shock ζ^L_{jt} to their planned labor. Again, this might represent optimization error or firm-specific transient frictions. For the sake of identification it is necessary for ζ^I and ζ^L to be

uncorrelated, because of the dependency of $\mathbb{L}$ on $k^*_{j,t+1}$. Due to the independence of these terms from the firm's information set, they are also independent of ζ^K_j. I require labor to be fully determined in the previous period so that it can be isolated from the productivity innovation η; this assumption is more at home in tightly regulated labor markets.[14]

Assumptions 10(i)-(iii) accommodate many possible input demand functions. Because prices are common among firms, allowing the input functions $\mathbb{G}, \mathbb{I}$, and $\mathbb{L}$ to depend on t implicitly allows them to depend on both present and past prices.[15] Dependence on lagged capital and labor permits the firm to consider adjustment costs as well. Further, because of the invertibility of $\mathbb{M}$, persistent productivity ω_{jt} depends only on prices and inputs in period t, so $\mathbb{G}$, $\mathbb{I}$ and $\mathbb{L}$ can be viewed as also depending on ω_{jt}, e.g., for any $\tilde{\mathbb{G}}(k^*, \ell, m, t, \omega)$,

$$\mathbb{G}(k^*_{jt}, \ell_{jt}, m_{jt}, t) = \tilde{\mathbb{G}}(k^*_{jt}, \ell_{jt}, m_{jt}, t, \mathbb{M}_t^{-1}(k^*_{jt}, \ell_{jt}, m_{jt}))$$

While I have assumed a balanced panel for notational simplicity, when firms do not enter simultaneously it must be clarified that the input functions depend on the year of the observation, not the age of the firm, while the 0 in the subscripts of $k^*_{j0}, \ell_{j0}, m_{j0}$ in Assumption 10(iv) refers to the period in which the firm enters.[16]

While the inclusion of an investment demand function $\mathbb{I}$ superficially resembles the approach of Olley and Pakes (1996), our investment demands play entirely different roles. Olley and Pakes's goal is to obtain a function which can be inverted for persistent productivity, while I use $\mathbb{I}$ (and $\mathbb{L}$) to break down each firm's time series of investment (and labor) into a sequence of decisions each depending only on a small number of variables. Since I never seek to invert $\mathbb{I}$ (or $\mathbb{L}$) for productivity, my estimator is robust to the well-known issues of zero investment and non-convex adjustment costs in the Olley-Pakes approach (see e.g., Levinsohn and Petrin (2003) for a discussion).

Assumption 10 may seem restrictive, but much of it is straightforward to relax. The choice in assumption 10(v) to use normal distributions is arbitrary; I approximate the integral in Equation (29) using Gauss-Hermite quadrature, and this approximation is simplified slightly when ζ^K_j is normally distributed. Somewhat more important is the additive separability of the residuals $\zeta^I, \zeta^L, \zeta^K$, which considerably simplifies the space of possible distributions. However, the primary convenience of Assumption 10 is the manner in which it restricts the interdependence between input variables in different periods, allowing $f(k^*_{j0}|m_{j0}, \mathbf{w_j})$ to be factored by Bayes' rule into a product of tractable density functions, one for each period.

[14] There are ways to generalize this assumption. For instance, if labor and intermediates are both determined statically I can instead follow an approach similar to GNR Appendix O6-3.

[15] If prices are heterogeneous across firms and observed, the functions $\mathbb{I}$ and $\mathbb{L}$ should be permitted to depend on price as well. In my empirical estimates I assume common prices due to data limitations, but my implementation permits observed prices.

[16] In principle it would be possible to condition input functions both on year of observation and year of entry, or equivalently, on year of observation and firm age. However, it would be conceptually meaningless to condition initial capital stock on both year of observation and firm age unless capital is not necessarily observed in the period in which the firm enters. As I discuss in Section 7.1, capital stocks are reported in only a few years in my data, but I sidestep this problem by relying on capital stock estimates derived from firm-reported depreciation, which are reported every year. That construction also introduces a possible relationship between initial capital and first-observed investment, so in the empirical application I allow $\mathbb{K}$ to condition on initial investment as well.

Equipped with Assumptions 4 and 10, the high-dimensional densities can be decomposed. The first decomposition is Equation (33) above; the second follows from Bayes' rule,

$$f_{k_0^*|m_0}(k_{j0}^*|m_{j0},\mathbf{w}_j) = \frac{f_{\varepsilon,0}(y_{j0},x_{j0}^*)f_{I,L}(I_{j1},\ell_{j1}|x_{j0}^*)f_{y,m|z^*}(y_{j1},m_{j1}|z_{j1}^*)f_{k^*|\ell,m}(k_{j0}^*|\ell_{j0},m_{j0})}{f_{w|\ell_0,m_0}(\mathbf{w}_j|\ell_{j0},m_{j0})} \tag{34}$$

The density in the denominator does not depend on k_{j0}^*, so it passes outside the integral in Equation (29). Multiplying both sides of (29) by $f_{w|\ell_0,m_0}$ turns the conditional likelihood of $(\mathbf{u}_j,k_{j0})$ given $\mathbf{w}_j$ into the joint likelihood of $(\mathbf{u}_j,\mathbf{w}_j,k_{j0})$ given only (ℓ_{j0},m_{j0}), which is the likelihood I maximize in my implementation, to avoid computing the marginal likelihood $f_{w|\ell_0,m_0}$.[17] Substituting both decompositions into Equation (29) and collecting terms in timing order gives

$$\begin{aligned} f_{u,w,k_0|\ell_0,m_0}(\mathbf{u}_j,\mathbf{w}_j,k_{j0}|\ell_{j0},m_{j0}) = \int & f_\xi(k_{j0}-k_0^*)f_{k^*|\ell,m}(k_0^*|\ell_{j0},m_{j0})f_{\varepsilon,0}(y_{j0},x_{j0}^*) \\ & \times \prod_{t=1}^{T-1} f_{I,L}(I_{jt},\ell_{jt}|x_{j,t-1}^*)f_{y,m|z^*}(y_{jt},m_{jt}|z_{jt}^*)dk_0^* \end{aligned} \tag{35}$$

where every k_{jt}^* inside the integral is the deterministic function of k_0^* and observed investment given by Assumption 4. This decomposition breaks down these extremely high-dimensional density functions into relatively low-dimensional densities. Then, the additional assumptions of additive separability and normality contained in Assumption 10 serve to further simplify the problem. Note also that the product now runs from $t=1$: the period-1 terms have exactly the same form as the later ones, so for each firm, every pair of consecutive periods contributes an $f_{y,m|z^*}$ factor.

It remains to write $f_{I,L}$ out. Because labor is chosen after investment is realized, the joint density factors in that order,

$$f_{I,L}(I_{jt},\ell_{jt}|x_{j,t-1}^*) = f_I(I_{jt}|x_{j,t-1}^*)f_L(\ell_{jt}|I_{jt},x_{j,t-1}^*)$$

where, writing $i_{jt}=\log I_{jt}$,

$$\begin{aligned} f_I(I_{jt}|x_{j,t-1}^*) &= 1-\mathbb{G}(x_{j,t-1}^*,t-1) \qquad \text{if } I_{jt}=0 \\ f_I(I_{jt}|x_{j,t-1}^*) &= \mathbb{G}(x_{j,t-1}^*,t-1)f_{\zeta^I}\left(i_{jt}-\mathbb{I}(x_{j,t-1}^*,t-1)\right) \qquad \text{if } I_{jt}>0 \\ f_L(\ell_{jt}|I_{jt},x_{j,t-1}^*) &= f_{\zeta^L}\left(\ell_{jt}-\mathbb{L}(k_{jt}^*,k_{j,t-1}^*,\ell_{j,t-1},m_{j,t-1},t-1)\right) \end{aligned}$$

In the last line k_{jt}^* is determined by $k_{j,t-1}^*$ and I_{jt} through Assumption 4; this is the channel through which labor responds to the realized investment shock. If labor is believed to be planned before observing ζ^I, the dependence on k_{jt}^* can be replaced with an indicator for whether $I_{jt}=0$, or dropped entirely if ℓ and I are planned simultaneously.

One further convenience of Assumption 10 is that it automatically implies Assumption 9, which Section 4 imposed without giving a sufficient condition. This fact is proved in Lemma 2.

[17] Theorem 1 still applies in this case. Suppose two different $\theta,\tilde{\theta}$ gave identical $f_{u,w,k_0|\ell_0,m_0}$ almost everywhere. Then they must also give identical $f_{w|\ell_0,m_0}$ almost everywhere, and therefore, identical $f_{u,k_0|w,\ell_0,m_0}$ almost everywhere that $f_{w|\ell_0,m_0}>0$. Since θ is identified by $f_{u,k_0|w,\ell_0,m_0}$, it is also identified by the more-detailed $f_{u,w,k_0|\ell_0,m_0}$.

Lemma 2. *Under Assumptions 1, 2, 3, 4, 8, and 10,*

$$f_{u|k,k^*,m_0}(\mathbf{u}_{j2}|k_{j0},k^*_{j0},\mathbf{w_j},m_{j0}) = f_{u|k^*}(\mathbf{u}_{j2}|k^*_{j0},\mathbf{w_j})$$

Proof. Begin by considering $T = 3$. The endogenous variables are given by

$$I_{j2} = \begin{cases} \exp\left(\mathbb{I}(x^*_{j1},1) + \zeta^I_{j2}\right) & \text{if the firm invests} \\ 0 & \text{otherwise} \end{cases}$$

$$\ell_{j2} = \mathbb{L}(k^*_{j2},k^*_{j1},\ell_{j1},m_{j1},1) + \zeta^L_{j2}$$
$$m_{j2} = \mathbb{M}_2(k^*_{j2},\ell_{j2},\omega_{j2})$$
$$y_{j2} = f(k^*_{j2},\ell_{j2},m_{j2}) + \omega_{j2} + \varepsilon_{j2}$$

where the investment occurs with probability $\mathbb{G}(x^*_{j1},1)$. Given k^*_{j0} and I_{j1}, k^*_{j1} is known by the perpetual inventory equation in Assumption 4, and k^*_{j2} is then determined by k^*_{j1} and I_{j2}. Then ω_{j1} can be constructed as

$$\omega_{j1} = \mathbb{M}_1^{-1}(k^*_{j1},\ell_{j1},m_{j1})$$

Recalling the productivity process given in Assumption 2, the endogenous variables depend only on $\zeta^I_{j2},\zeta^L_{j2},\eta_{j2},\varepsilon_{j2}$ and $(k^*_{j0},\mathbf{w_j})$. By the independence assumptions contained in Assumptions 2 and 10, each of these shocks is independent of $\mathcal{I}_{j1}$, and $m_{j0} \in \mathcal{I}_{j1}$, so m_{j0} can provide no additional information about the period-2 endogenous variables after conditioning on k^*_{j0} and $\mathbf{w_j}$. Intuitively, conditioning on the period 1 variables insulates the information on productivity contained in m_{j0} from m_{j2}. Independence from k_{j0} is immediate from Assumption 8(i).

For $T > 3$ the argument proceeds by induction on t. Suppose x^*_{jt} is a function of $(k^*_{j0},\mathbf{w}_j)$ and of shocks dated 2 through t that are independent of $\mathcal{I}_{j1}$. The equations above, advanced one period, write the period $t+1$ endogenous variables as a functions of x^*_{jt}, ω_{jt}, and shocks (independent of m_{j0} by assumption) dated $t+1$; since $\omega_{jt} = \mathbb{M}_t^{-1}(x^*_{jt})$, m_{j0} can provide no additional information about period t endogenous variables after conditioning on k^*_{j0} and w_j. □

5.2 Sieves

With the likelihood factored into low-dimensional pieces, it remains to write it in terms of sieves. In order to estimate Equation (29) semiparametrically, I use a sieve maximum likelihood estimator. Following HS, my estimator, which I call MLE-HS, is

$$\hat{\theta} = \underset{\theta\in\mathcal{A}_N}{\operatorname{argmax}} \frac{1}{N}\sum_{j=1}^{N} \log f_{u,w,k_0|\ell_0,m_0}(\mathbf{u}_j,\mathbf{w}_j,k_{j0}|\ell_{j0},m_{j0};\theta) \tag{36}$$

Here θ is a (possibly infinite-dimensional) vector defining the densities $f_{y,m|z^*}$, f_ξ, $f_{k^*|\ell,m}$. This includes the parameters of the functions f and h, the densities of ε, η, ζ^I, ζ^L and ζ^K, and the mean functions $\mathbb{K}$, $\mathbb{L}$, $\mathbb{I}$ and $\mathbb{G}$. $\{\mathcal{A}_N\}$ is a sequence of sieve spaces that grows with the sample size N.

The unknown mean functions $\mathbb{K}$, $\mathbb{L}$, $\mathbb{I}$, $\mathbb{G}$, and the productivity process h are estimated as polynomial tensor product sieves. For instance,

$$\mathbb{K}(\ell_{j0}, m_{j0}, t_{j0}) = \sum_{p=0}^{D_{KL}} \sum_{q=0}^{D_{KM}} \sum_{r=0}^{D_{KT}} \theta^K_{pqr} \ell^p_{j0} m^q_{j0} t^r_{j0}$$

where $D_K = (D_{KL}, D_{KM}, D_{KT})$ are hyperparameters controlling the dimension of the sieve.[18] Likewise, $\mathbb{L}$ is controlled by $D_L = (D_{LK}, D_{LK'}, D_{LL}, D_{LM}, D_{LT})$; I parameterize $\mathbb{L}$ in terms of $(k^*_{j,t-1}, I_{jt}, \ell_{j,t-1}, m_{j,t-1}, t)$ rather than $(k^*_{jt}, k^*_{j,t-1}, \dots)$, which is equivalent by Assumption 4. Both $\mathbb{I}, \mathbb{G}$ are controlled by the shared hyperparameters $D_I = (D_{IK}, D_{IL}, D_{IM}, D_{IT})$. To ensure that $\mathbb{G}$ stays within the desired range, the normal CDF is applied to the tensor product sieve:

$$\mathbb{G}(k^*_{jt}, \ell_{jt}, m_{jt}, t) = \Phi\left(\sum_{p=0}^{D_{IK}} \sum_{q=0}^{D_{IL}} \sum_{r=0}^{D_{IM}} \sum_{s=0}^{D_{IT}} \theta^G_{pqrs} (k^*_{jt})^p \ell^q_{jt} m^r_{jt} t^s \right)$$

I can also include time fixed effects in the mean functions, for instance,

$$\mathbb{K}(\ell_{j0}, m_{j0}, t_{j0}) = \sum_{p=0}^{D_{KL}} \sum_{q=0}^{D_{KM}} \sum_{r=0}^{D_{KT}} \theta^K_{pqr} \ell^p_{j0} m^q_{j0} t^r_{j0} + \theta^K_t$$

with $\theta^K_{0,0,1} \dots \theta^K_{0,0,D_{KT}} = 0$ to avoid collinearity. I include time fixed effects in both my Monte Carlo and empirical estimates.

The densities of the residuals ε, η, $\xi_j = k_{j0} - k^*_{j0}$, ζ^I, ζ^L, and ζ^K can each be modeled as mixtures of normals, possibly time-varying, constrained to have zero mean, with the number of mixture components growing with the sample size. However, in my empirical and Monte Carlo estimation, I model each residual as having a time-independent normal distribution. This is numerically equivalent to using a normal mixture with one component, or a first-order Edgeworth expansion.

Similarly, the productivity process $h(\cdot)$ and the functions $\mathcal{C}$ and $D^{\mathcal{E}}$ making up the production function f are also polynomial sieves. In my main estimates, I use a translog production function for f with a degree 3 polynomial for h; this corresponds to a degree-2 tensor product sieve for f with terms above degree two restricted to zero. Because Cobb-Douglas technology and AR(1) productivity are both widespread in empirical work, I also present a set of empirical results corresponding to those functional forms.

Consequently, the sieve space $\mathcal{A}_N$ can be characterized by the hyperparameters (D_K, D_L, D_I). While determining the optimal rate of growth for the dimension of $\mathcal{A}_N$ is possible (see, e.g., Chen, Christensen, and Kankanala, 2024, finding optimal sieve dimension for the nonparametric IV estimator of W. Newey and Powell, 2003), I do not do so here. In both my empirical and Monte Carlo estimates I set each sieve to be of dimension 0 with respect to

[18] Since my panel is unbalanced, period 0 is in different years for different firms. I allow initial capital to depend on year of first observation t_{j0} as average capital endowment of new firms may not be constant over time.

time ($D_{KT} = D_{LT} = D_{IT} = 0$), and of dimension 1 for each other variable and include time fixed effects.[19]

Several additional technical assumptions are necessary to assure consistency of the sieve maximum likelihood estimator; see HS Appendix 1 for an overview.

6 Monte Carlo

I first test my estimator on simulated Monte Carlo data. A full description of the data-generating process is given in Appendix A; I briefly summarize it here. A balanced panel of 600 firms is observed over 6 periods with measurement error in initial capital. Firms have a Cobb-Douglas production function and an AR(1) process for the persistent component of productivity:

$$f(k, \ell, m) = \alpha_K k + \alpha_L \ell + \alpha_M m$$
$$h(\omega) = \gamma_0 + \gamma_1 \omega$$

but this is not imposed at estimation time. Firms maximize the expected sum of discounted profits. Investment, labor, and intermediates are chosen by solving the firm's problem. Investment and labor are chosen one period before they are realized, while intermediates are assumed to follow a static process.

To obtain an approximate steady-state distribution, firms are simulated for 100 periods, but only the final 6 periods are used for each firm. Normally distributed measurement error is added in the first observed period and initially enters classically. In subsequent periods the distribution of measurement error becomes non-classical due to firm investment.

To evaluate the robustness of my estimator I consider several additional Monte Carlo data-generating processes. While my estimator does not in principle require strong assumptions on the distributions of the various residuals, in my implementation I treat the residuals $\varepsilon, \eta, \zeta^I, \zeta^L, \zeta^K, \xi$ as normally distributed, so my estimator may be sensitive to the distributions of residuals. The first two of my alternative DGPs allow ε and η, respectively, to be drawn from fat-tailed distributions. I find that my estimator performs well under the alternative specifications for η and ε, while standard methods perform no better (and in the case of fat-tailed η, much worse) than in the baseline DGP.

While I have focused on the case of initial measurement error, in reality, error may enter in multiple ways. Because Assumption 4 is crucial for the identification of my estimator, there is no formal guarantee that my estimator will even correct for initial measurement error when additional sources of error contaminate the sample. To investigate the robustness of my estimator to additional sources of measurement error, I consider several additional data-generating processes. These include a misspecified law of motion for capital, random depreciation rates, and measurement error in observed investment.

6.1 Monte Carlo Results

I estimate my Monte Carlos using the GNR estimator (which I label GMM), MLE, GMM-HS, and MLE-HS.

[19] This is inspired by the common choice in ACF estimation to use a time fixed effect instead of continuous time terms.

Table 1 gives Monte Carlo estimates of the mean of the elasticities of capital and labor and mean returns to scale, along with the standard deviation of log productivity $\omega + \varepsilon$. True productivity is estimated by applying the GNR estimator to the error-free data, while the production function is known and common to all firms.

When error-uncorrected estimators are used, measurement error causes downward bias in capital elasticity as the correlation between observed capital and output is attenuated by measurement error. The elasticity of labor rises to compensate. However, the increase in labor elasticity does not completely cancel out the decrease in capital elasticity and returns to scale are slightly underestimated. Since less of the variation in output can be explained by the variation in observed inputs, the variance of productivity is overestimated. By contrast, the error-corrected GMM-HS and MLE-HS estimators produce accurate and precise estimates of each of the parameters and moments under consideration.

In Tables 5, 6, and 7 in Appendix C.1, I give Monte Carlo estimates under alternative data-generating processes. While the alternative data-generating processes are described formally in Appendix A, I give an overview here.

Table 5 relaxes the functional-form and distributional assumptions of the baseline DGP. Although the distributions of the shocks $\varepsilon_{jt}, \eta_{jt}$ are nonparametrically identified, my implementation takes both shocks to be normal, so a natural concern is that the estimator's performance depends on those choices being close to correct. The first panel reproduces the baseline estimates for comparison; the second and third replace the normal transient shock ε and the normal innovation η, respectively, with Laplace random variables of the same variance. In each case, the error-corrected estimators remain close to the true values (though less precisely estimated in the Laplace η panel) suggesting that a close match between the choice of sieves for f_ε, f_η and the true distributions is not required for accurate recovery of production functions and productivity. However, while the DGP with Laplace shock ε shows similar biases to the baseline case when error-correction is neglected, the DGP with Laplace innovation η shows far more bias in elasticities of capital and labor, but less bias in returns to scale and productivity distribution.

Tables 6 and 7 instead violate Assumption 4, which is what identifies my estimator, so there is no formal guarantee that it corrects even the initial measurement error under these processes, not even asymptotically. In Table 6 the true depreciation rate varies across firms (i.i.d. over time in the second panel, constant over time for each firm in the third) while the econometrician constructs capital using the average rate. Table 7 considers two more ways the constructed series can go wrong for reasons unrelated to the initial stock: noise in the econometrician's observation of investment in each period, and an assumed depreciation rate that differs from the truth. As before, the first panel of each table gives results under the baseline DGP, for ease of comparison. The error-corrected estimators continue to reduce bias in elasticity substantially in all four cases. However, when investment is measured with error, the MLE-HS estimator (but not the GMM-HS estimator) modestly increases the upward bias in estimated productivity standard deviation. I would not read these results as showing that my estimator handles arbitrary sources of measurement error; rather, they suggest that my estimator is able to correct the initial-stock channel even when other sources of error are present.

Table 1: Monte Carlo Estimates

		Baseline Model			
	Truth	GMM	GMM-HS	MLE	MLE-HS
Elasticities					
Capital	0.20	0.11	0.20	0.13	0.20
		(0.01)	(0.02)	(0.05)	(0.01)
Labor	0.30	0.33	0.30	0.33	0.30
		(0.01)	(0.01)	(0.06)	(0.01)
RTS	1.00	0.93	1.01	0.96	1.00
		(0.01)	(0.01)	(0.02)	(0.01)
Productivity					
S. D.	0.61	0.69	0.60	0.74	0.61
	(0.01)	(0.02)	(0.02)	(0.09)	(0.01)

Comparison of Monte Carlo estimates of log productivity and elasticity distributions to true values. The DGP is described in Appendix A. The first column gives the average standard deviation of true productivity across simulations, the true values of capital and labor elasticities, and the true returns to scale (RTS); I simulate data assuming a Cobb-Douglas production function, so that the true elasticities and RTS are constant across firms. Four estimators are applied: error-uncorrected GMM and MLE, and error-corrected GMM-HS and MLE-HS. In each case I estimate a translog production function and a degree 3 productivity process. In each simulation, the mean of each elasticity and returns to scale, and the standard deviation of productivity, are calculated for each estimator; I report the average of these statistics across all 200 simulations, with Monte Carlo standard deviations in parentheses.

7 Empirical Estimates

Having shown that my estimator recovers the true parameters in simulated data, I now apply it to Chilean manufacturing data.

7.1 Data

I apply my estimators to data drawn from the widely-used *Encuesta Nacional Industrial Anual* (ENIA), a high-quality panel census of all manufacturing firms in Chile with over 10 employees; my dataset covers every year from 1979 to 1996. Greenstreet (2007) provides a summary, and Liu (1991) provides a more detailed description. This panel has been widely used in the production function literature (e.g., Levinsohn and Petrin, 2003). Following GNR, I estimate separately on each of the five largest industries at the three-digit level. In my main specification I use a translog production function and a cubic polynomial for the productivity process h.

For labor I use number of employees weighted by wage, following GNR. Because the survey only reports capital in 6 years (out of 18), and because of concerns about data quality, Greenstreet constructs an alternative measure of capital, using reported depreciation to obtain an approximate starting point (a k_{j0}) and perpetual inventory to construct the rest of the series. I follow Greenstreet's approach to generate my initial capital estimates. To avoid estimating multiple mismeasured stocks of capital, in the perpetual inventory equation I follow the common practice of aggregating across capital types and assuming a single depreciation rate of $\delta = 0.1$. However, it would be straightforward to extend my estimator to handle vector-valued capital stocks, as long as reported capital and initial intermediates (or other instruments) are also a vector with sufficient variation to satisfy the nonparametric rank condition in Assumption 6. This implies there must be at least as many instruments as varieties of true capital.

Greenstreet's capital estimates include a correction term justified under an assumption that firm investment takes place in continuous time. Greenstreet's approach introduces a possible relationship between initial capital and investment reported in the first observed period, so I allow $\mathbb{K}$ to depend on initial investment as well as labor and intermediates.

7.2 Empirical Results

In Table 2 I give estimates of moments of the distributions of elasticities, capital, and productivity ($\varepsilon_{jt} + \omega_{jt}$) for the five largest 3-digit industries. The column GMM corresponds to the standard GNR estimator. In all five industries, the effect of error correction is directionally similar to my Monte Carlo simulations: higher capital elasticity, lower labor elasticity, and lower productivity variance. Capital elasticity is underestimated by up to 49.9% and labor elasticity is overestimated by up to 39.5%; the average percentage biases across industries are 38.7% and 18.9%, respectively. The proportion of productivity variance attributable to measurement error ranges from 12.2% to 31.2% (22.7% on average across industries). Heterogeneity in capital elasticity is also frequently underestimated, most severely in industry 321, where the standard deviation of elasticity is underestimated by 71.0%. Variance in the elasticity of other inputs is also sometimes severely underestimated: the S.D. of labor elasticity

is underestimated by 42% in industry 321 and the S.D. of intermediate elasticity is underestimated by 25% in industry 311. The translog specification does not restrict elasticity to be positive, but despite the large elasticity variances in some industries, negative elasticities were rare, with no industry having a negative estimated capital elasticity in more than 5% of observations or a negative estimated labor elasticity in more than 4% of observations.

Table 2: Empirical Estimates

		Industry									
		311		321		322		331		381	
		GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS
Elasticities											
Capital	Mean	0.12	0.15	0.11	0.18	0.10	0.19	0.07	0.13	0.15	0.25
	S.D.	0.05	0.06	0.03	0.11	0.03	0.05	0.03	0.05	0.05	0.06
Labor	Mean	0.28	0.28	0.43	0.33	0.40	0.36	0.42	0.36	0.51	0.36
	S.D.	0.09	0.09	0.11	0.19	0.09	0.09	0.11	0.10	0.10	0.08
Interm.	Mean	0.66	0.64	0.54	0.54	0.55	0.54	0.58	0.56	0.49	0.50
	S.D.	0.09	0.12	0.11	0.12	0.10	0.12	0.10	0.11	0.09	0.13
Total	Mean	1.05	1.06	1.08	1.05	1.05	1.09	1.06	1.05	1.15	1.11
	S.D.	0.04	0.03	0.04	0.04	0.02	0.03	0.04	0.04	0.04	0.03
Capital											
	Var	4.47	3.27	3.43	2.78	2.29	1.86	3.46	3.13	3.41	2.86
Productivity											
	Var	0.09	0.08	0.13	0.10	0.11	0.08	0.15	0.12	0.15	0.10
% Bias			12.27		18.59		30.80		20.78		31.22
Firms		2739		967		972		957		1241	
Periods		18		18		18		18		18	
Firm-Period Pairs		20849		5879		5143		5432		6262	

Empirical estimates for the largest 3-digit industries in the sample. Two estimators are applied: error uncorrected GMM and error-corrected GMM-HS. In each case I estimate a translog production function and a degree 3 productivity process. The mean and standard deviation of each elasticity, the standard deviation of productivity, and the percentage of productivity variance attributable to measurement error are computed for each industry.

Table 3: Empirical Estimates: Cobb-Douglas

	Industry									
	311		321		322		331		381	
	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS
Elasticities										
Capital	0.16	0.16	0.13	0.17	0.12	0.16	0.09	0.11	0.16	0.18
Labor	0.21	0.20	0.36	0.32	0.36	0.33	0.39	0.37	0.44	0.42
Interm.	0.64	0.64	0.51	0.51	0.50	0.50	0.56	0.56	0.46	0.46
Total	1.01	1.01	1.00	1.00	0.98	0.99	1.04	1.03	1.07	1.07
Capital										
Var	4.47	3.81	3.43	2.59	2.29	1.79	3.46	2.82	3.41	2.61
Productivity										
Var	0.11	0.11	0.14	0.14	0.13	0.12	0.16	0.16	0.16	0.15
% Bias		4.68		2.21		4.90		0.95		6.34
Firms	2739		967		972		957		1241	
Periods	18		18		18		18		18	
Firm-Period Pairs	20849		5879		5143		5432		6262	

Empirical estimates for the largest 3-digit industries in the sample. Two estimators are applied: error uncorrected GMM and error-corrected GMM-HS. In each case I estimate a Cobb-Douglas and an AR(1) productivity process. The mean and standard deviation of each elasticity, the standard deviation of productivity, and the percentage of productivity attributable to measurement error are computed for each industry.

Table 4: Heterogeneity in Productivity

	Industry									
	311		321		322		331		381	
	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS	GMM	GMM-HS
75-25 ratio	1.39	1.40	1.50	1.44	1.44	1.37	1.51	1.47	1.56	1.46
90-10 ratio	1.93	1.91	2.20	2.06	2.16	1.89	2.38	2.18	2.40	2.13
95-5 ratio	2.51	2.36	2.92	2.66	2.86	2.40	3.19	2.86	3.22	2.72
Exporter	1.49	11.47	-0.52	0.83	9.68	3.10	0.38	2.09	-4.61	-4.83
Importer	13.16	24.68	7.79	6.59	13.66	5.50	17.33	11.95	8.89	0.49
Advertiser	3.58	7.79	2.53	3.57	6.61	2.88	3.30	3.03	0.32	-0.60
Wage > median	24.86	32.16	18.73	12.54	22.68	12.50	24.71	19.19	21.78	8.65

Impact of capital measurement error on estimated productivity heterogeneity. Empirical estimates for the largest 3-digit industries in the sample. Two estimators are applied: error uncorrected GMM and error-corrected GMM-HS. In each case I estimate a translog production function and a degree 3 productivity process. The first three rows report ratios of productivity for plants at different percentiles of estimated productivity. The following four rows contain the percentage difference in average productivity between firms with differing characteristics.

In Table 3 I give corresponding estimates under a Cobb-Douglas production function. While productivity and the production function are biased more modestly, the impact of measurement error is directionally similar. Nevertheless, even with a linear production function, AR(1) productivity process, and relatively long panel, I obtain average biases of 15.1% in capital elasticity and 7.4% in labor elasticity, and an average of 3.8% of productivity variation is attributable to measurement error.

One interesting empirical regularity is that in every industry, the uncorrected Cobb-Douglas/AR(1) capital elasticity estimates are closer to the corresponding corrected average elasticity estimates in the translog model. For instance, in industry 311, the average capital elasticity estimated by GMM-HS (0.15) is closer to the uncorrected Cobb-Douglas/AR(1) estimate (0.16) than to the uncorrected translog estimate (0.12). The same is usually true of labor elasticity, except in industry 311. This underscores the tendency of measurement error to more badly bias estimates of nonlinear production functions.

In Table 4 I report my estimates of several statistics of the distribution of $\exp(\omega_{jt} + \varepsilon_{jt})$ which are of empirical interest. The first three rows give the ratios of percentiles of the productivity distribution; for instance, in industry 311 (food products), I estimate that a firm at the 75th percentile of productivity will produce 40% more output than a firm at the 25th percentile, given equal inputs. The next four rows show the relationship between productivity and observable firm characteristics; for instance, for industry 311, I estimate that firms which export produce on average 11.47% more output than non-exporters, given the same inputs. The impact on correlations between productivity and observables is often large, up to 13.1 percentage points. Unlike mean elasticity estimates, the bias in these correlations is of ambiguous sign, suggesting that error-naive estimates cannot be interpreted as bounds.

The GMM-HS estimates reveal that error-uncorrected methods often substantially overstate the correlation between productivity and importer status or high wages. The correlation with advertiser status also appears to generally be overstated, but is already quite small. In 321, the correlation between exporter status and productivity changes sign and becomes positive after error-correction, although again, the correlations involved are quite small.

If, as I assume, measurement error enters mainly through the initial stock, it should be expected to decline over time as the measurement error in initial capital stock depreciates and is replaced by correctly-measured observed investment. To test this, I created subpanels for each industry by truncating at different end-years, and estimated under the Cobb-Douglas assumptionfor each subpanel. Figure 1 shows that as the panel length increases, the GMM estimates of capital elasticity become more accurate. This implies that bias due to initial measurement error is likely to be more severe in shorter panels. However, convergence in bias is relatively slow except in the largest industry (Food Products, SIC code 311). One reason for this may be that since the panels are unbalanced, new firms with new, large measurement errors continually enter, while old firms with severely decayed measurement errors continually exit; in the full dataset (all industries), the average firm operates for only six years.

8 Conclusion

Despite repeated concerns in the literature about the potential impact of measurement error in capital on production functions and productivity estimates, no approach to correcting

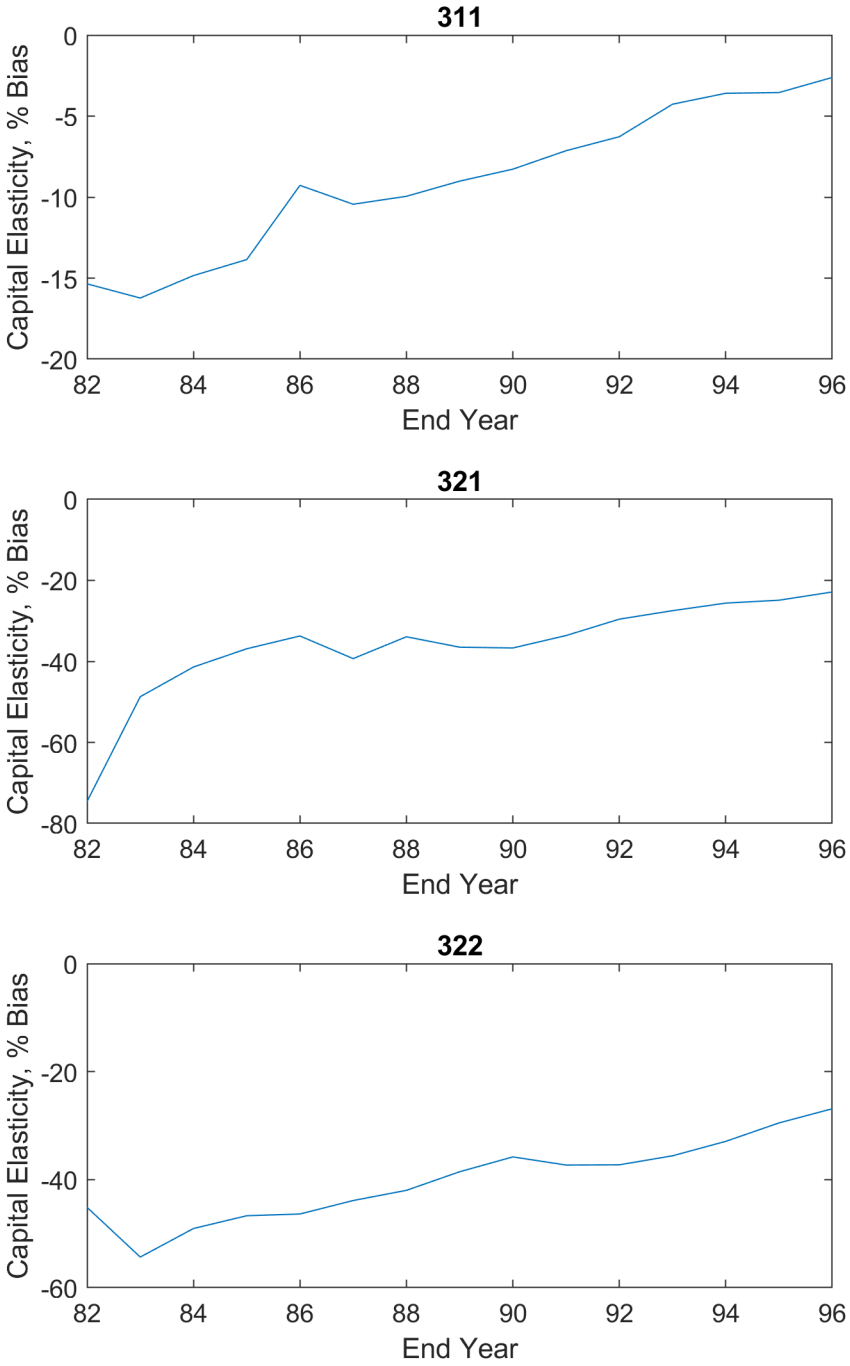

Figure 1: Graph of percentage difference between GMM and GMM-HS estimates of capital elasticity for panels truncated in different years. The value plotted for each year is the percentage difference between the GMM and GMM-HS estimates of capital elasticity for the subpanel truncated in that year for the specified industry. In each case I estimate a Cobb-Douglas production function and an AR(1) productivity process.

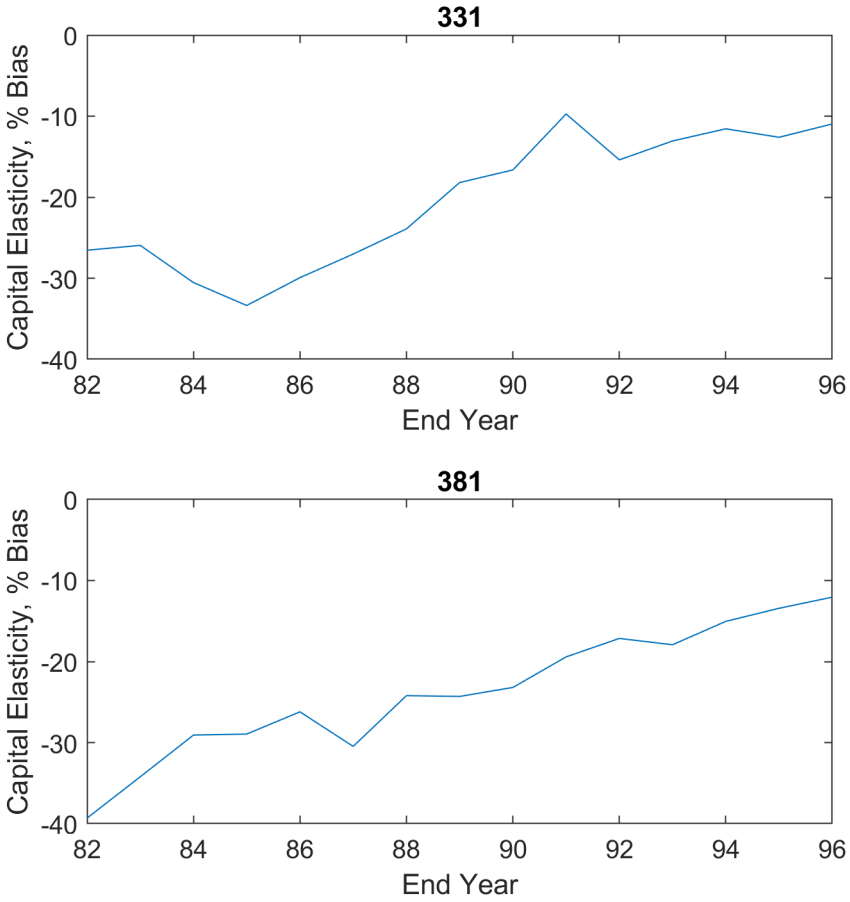

Figure 1: (cont'd) Graph of percentage difference between GMM and GMM-HS estimates of capital elasticity for panels truncated in different years. The value plotted for each year is the percentage difference between the GMM and GMM-HS estimates of capital elasticity for the subpanel truncated in that year for the specified industry. In each case I estimate a Cobb-Douglas production function and an AR(1) productivity process.

for capital mismeasurement has become widespread. Focusing on the issue of measurement error in initial capital stock, I construct a production function estimator based on a popular method but which corrects for measurement error which is potentially made non-classical by the typical perpetual inventory approach to the construction of capital series. This estimator is identified by using the structural information provided by the perpetual inventory equation to exploit standard assumptions on the timing of input decisions, allowing intermediate inputs to be used as an instrument for mismeasured capital.

I implement my estimator and apply it to both Monte Carlo simulations and Chilean manufacturing data. My Monte Carlo simulations show that my estimator can correctly recover production functions and productivity distributions when initial capital is measured with error. While the distributional assumptions used for tractability place strong restrictions on the data-generating process, I also show by Monte Carlo that my estimator can perform well when distributional assumptions are violated and even when measurement error enters through channels other than initial capital.

My empirical estimates suggest that failing to correct for measurement error can result in severely biased estimates of production functions, the dispersion of productivity, and correlations between productivity and other observables. This bias is most severe in nonlinear models, but decreases in panels of increasing length. However, even in relatively lengthy panels (18 periods) the bias on estimated productivity may be as large as 31%.

A Monte Carlo Details

In this Appendix, I give details of my Monte Carlo simulations. For my baseline estimates I use a panel of $N = 600$ firms (indexed by j) over $T = 6$ periods. In order to draw from an approximate steady-state distribution, for each firm I simulate 100 periods and discard the first 94. Except as specified below, all parameters are chosen to approximately match the corresponding values in the full sample.

A.1 Production Function and Productivity

Firm log output is given by

$$y_{jt} = \alpha_K k^*_{jt} + \alpha_L \ell_{jt} + \alpha_M m_{jt} + \omega_{jt} + \varepsilon_{jt}$$

with $\alpha_K = 0.20$, $\alpha_L = 0.30$, $\alpha_M = 0.5$. Productivity ω is drawn from a stationary distribution following the AR(1) process:

$$\omega_{j,t+1} = \gamma_0 + \gamma_1 \omega_{jt} + \eta_{j,t+1}$$

where $\gamma_0 = 0.1$, $\gamma_1 = 0.8$ and $\eta_{jt} \sim N\left(0, 0.25\sqrt{1-\gamma_1^2}\right)$ so that the standard deviation of ω_{jt} is 0.25. ε is drawn iid across firms and time from a normal distribution with mean zero and variance $\sigma^2_\varepsilon = 0.45$.

Under Alternative DGP 1, the productivity process is a cubic function:

$$\omega_{j,t+1} = \gamma_0 + \gamma_1 \omega_{jt} + \gamma_2 \omega^2_{jt} + \gamma_3 \omega^3_{jt} + \eta_{j,t+1}$$

where $(\gamma_0, \gamma_1, \gamma_2, \gamma_3) = (0.1, 0.8, -0.1, -0.8)$, and $\eta \sim \text{Uniform}(-0.1, 0.1)$. γ_2, γ_3 and the distribution of η are chosen to ensure that ω_{jt} has a steady-state distribution: $h(\omega) \in [-0.9, 0.9]$ for $\omega \in (-1, 1)$, so $\omega_{j,t+1} \in [-1, 1]$ for $\omega_{jt} \in [-1, 1]$. This results in a lower standard deviation for η, approximately 0.05 compared to approximately 0.15 under the baseline DGP.

Under Alternative DGP 2, η is drawn from a Laplace distribution with scale parameter $b = \sigma_\eta/\sqrt{2}$, resulting in the same variance of η as in the baseline DGP. Similarly, under Alternative DGP 3, ε is drawn from a Laplace distribution with scale parameter $b = \sigma_\varepsilon/\sqrt{2}$.

A.2 Profit and Prices

In each period, the firm's profit function is

$$\pi_{jt} = P_t^Y Y_{jt} - \frac{c}{2} I_{jt}^2 - P_t^L L_{jt} - P_t^M M_{jt}$$

In each period t, the firm chooses inputs so as to maximize its expected sum of profits $\sum_{\tau=t}^{\infty} \beta^{\tau-t} \mathrm{E}_t[\pi_{j\tau}]$. I set $c = 0.5$ and $P_t^M = 1$ in all periods, while the log prices of output and labor each follow an AR(1) process:

$$p_{t+1}^Y = \theta_0^Y + \theta_1^Y p_t^Y + \zeta_{t+1}^{P_Y}$$
$$p_{t+1}^L = \theta_0^L + \theta_1^L p_t^L + \zeta_{t+1}^{P_L}$$

with normally distributed ζ^{P_Y}, ζ^{P_L}. Firms face common prices for labor and output. I set $\theta_1^Y = \theta_1^L = 0.3$, $\theta_0^Y = 0$ and $\theta_0^L = -1.5$. ζ^{P_Y} has standard deviation 0.2 while ζ^{P_L} is chosen so that the standard deviation of labor price will be 0.01.[20]

A.3 Firm Input Choices

In period t, the firm observes current prices p_t^Y and p_t^L and its productivity ω_{jt}. It chooses a level of investment $\overline{I}_{j,t+1}$ and draws a shock $\zeta_{j,t+1}^I \sim N(0, \sqrt{7})$ so that, in period $t+1$, its realized investment will be $I_{j,t+1} = \overline{I}_{j,t+1} \exp(\zeta_{j,t+1}^I)$. After observing ζ_{jt}^I, it chooses $\overline{\ell}_{j,t+1}$ and draws another shock $\zeta_{j,t+1}^L \sim N(0, \sqrt{0.45})$, so that in period $t+1$ its realized log labor will be $\ell_{j,t+1} = \overline{\ell}_{j,t+1} + \zeta_{j,t+1}^L$.

Under the above assumptions, I can adapt the approach of Syverson (2001) to obtain a closed form for the firm's choice of ℓ_{jt} and a tractable series solution for the firm's choice of I_{jt}. Similar approaches are used in Biesebroeck (2003) and Ackerberg, Caves, and Frazer (2015). Specifically,

$$\bar{\ell}_t = \frac{1-\alpha_M}{\alpha_K} \left(\log \left(\alpha_L \alpha_M^{\frac{\alpha_M}{1-\alpha_M}} \mathcal{E}^{\frac{1}{1-\alpha_M}} \mathcal{Z}_L^{-\frac{\alpha_K}{1-\alpha_M}} \frac{\mathrm{E}_{t-1}\left[\left(P_t^Y \exp(\omega_t) \right)^{\frac{1}{1-\alpha_M}} \right]}{\mathrm{E}_{t-1}\left[P_t^L \right]} \right) \right) + k_t^*$$

$$\overline{I}_{j,t+1} = \frac{\alpha_K \alpha_L^{\frac{\alpha_L}{\alpha_K}} \alpha_M^{\frac{\alpha_M}{\alpha_K}} \mathcal{E}^{\frac{1}{\alpha_K}} \mathcal{Z}_L^{\frac{\alpha_L}{1-\alpha_M}} \beta}{c} \sum_{\tau=1}^{\infty} \left[\left(\frac{(1-\delta)\beta}{\mathcal{Z}_I} \right)^{\tau-1} \left(\frac{\mathrm{E}_t\left[\left(P_{t+\tau}^Y \exp\left(\omega_{j,t+\tau} \right) \right)^{\frac{1}{1-\alpha_M}} \right]}{\mathrm{E}_t\left[P_{t+\tau}^L \right]} \right)^{\frac{1-\alpha_M}{\alpha_K}} \right]$$

[20] This is lower than the observed price variation in the full sample, to help approximately match the labor variance observed in the data.

where

$$\mathcal{Z}_L = \mathrm{E}_{t-1}[\exp(\zeta_{jt}^L)]$$

$$\mathcal{Z}_I = \mathrm{E}_{t-1}[\exp(\zeta_{jt}^I)]$$

I approximate the infinite sum by truncating after 100 periods. Because prices and productivity follow AR(1) processes, and productivity is independent of prices, all of the expectations are easy to evaluate in the standard DGP. Under Alternative DGP 1, when h is cubic, I approximate the expected value of future productivity $\mathrm{E}[\exp(\omega_{j,t+\tau})]$ by Monte Carlo simulation.

A.4 Measurement Error

In my baseline DGP, measurement error enters initially

$$K_{j0} = K_{j0}^* \exp(\xi_{j0})$$

Observed capital is constructed according to perpetual inventory:

$$K_{j,t+1} = (1-\delta)K_{jt} + I_{j,t+1}$$

following the same law of motion as true capital. In Alternative DGP 4, capital is constructed using an incorrect law of motion:

$$K_{j,t+1} = (1-\tilde{\delta})K_{jt} + I_{j,t+1}$$

where $\tilde{\delta} = 0.15$ or 0.2. The true depreciation rate remains 0.1.

In Alternative DGP 5 and 6, the average depreciation rate is known, but depreciation is not constant. Instead, the true depreciation is drawn from a beta distribution chosen to have a mean of 0.1 and a standard deviation of either 0.05 or 0.20. In DGP 5, depreciation is i.i.d. across firms and time:

$$K_{j,t+1}^* = (1-\delta_{jt})K_{jt}^* + I_{j,t+1}$$

while in DGP 6 it is i.i.d. across firms and constant across time:

$$K_{j,t+1}^* = (1-\delta_j)K_{jt}^* + I_{j,t+1}$$

In DGP 5, to simplify the firm's dynamic programming problem, firms know their entire future series of δ_{jt}. In either DGP, capital is constructed using the average depreciation rate:

$$K_{j,t+1} = (1-\mathrm{E}[\delta_{jt}])K_{jt} + I_{j,t+1}$$

In Alternative DGP 7, investment is measured with error ξ_{jt}^I. Observed capital is constructed according to this mismeasured investment:

$$K_{j,t+1} = (1-\delta)K_{jt} + I_{j,t+1}\exp(\xi_{j,t+1}^I)$$

while true capital is constructed as in the baseline DGP.

A.5 Monte Carlo Sieve

For my main Monte Carlo results, I choose a sieve with the following dimensions:

$$D_{KM} = D_{KL} = 1$$

$$D_{LK} = D_{LL} = D_{LM} = D_{LK'} = 1$$

$$D_{IK} = D_{IL} = D_{IM} = 1$$

$$D_{KT} = D_{LT} = D_{IT} = 0$$

Time fixed effects are used in estimating $\mathbb{L}$ and $\mathbb{I}$. This is the same sieve I use in my empirical estimates, except that I do not include time fixed effects in $\mathbb{K}$ or consider the possibility of zero investment, because all firms enter simultaneously and invest in each period. Because all firms are first observed in the same period, setting $D_{KT} = 0$ and omitting fixed effects in $\mathbb{K}$ is of no consequence, but $D_{LT} = D_{IT} = 0$ means that the estimator cannot take intertemporal variation in labor prices into account except through fixed effects when estimating $\mathbb{L}$ and $\mathbb{I}$. Because the true distribution of $\overline{I}$ described in the previous section depends nonlinearly on observables, this sieve cannot precisely match it, which speaks to the robustness of my estimator.

B Proofs

B.1 Proof of Lemma 1

Begin by considering the density of ε. At θ^0, $f_{\varepsilon,t}$ evaluated at output and inputs for a given observation gives the true density of the transient shock at that point.

$$f^0_\varepsilon(\varepsilon_{jt}) = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0) \tag{37}$$

To relate $f_{\varepsilon,t}$, defined in the statement of Lemma 1, to the density of y, consider $f_{\varepsilon|x^*,x^*_{-1}}$ and $f_{y|x^*,x^*_{-1}}$, the true densities of ε and y (respectively) conditional on current and lagged inputs. A trivial change of variables[21] establishes that

$$f_{y|x^*,x^*_{-1}}(y_{jt}|x^*_{jt}, x^*_{j,t-1}) = f_{\varepsilon|x^*,x^*_{-1}}(\tilde{\varepsilon}_t(y_{jt}, x^*_{jt}; \theta^0)|x^*_{jt}, x^*_{j,t-1})$$

But by Assumption 2, the density of ε is independent of inputs, so

$$f_{y|x^*,x^*_{-1}}(y_{jt}|x^*_{jt}, x^*_{j,t-1}) = f^0_\varepsilon(\tilde{\varepsilon}_t(y_{jt}, x^*_{jt}; \theta^0)) = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0)$$

[21] Substitute Equation 4 into 7 and observe that

$$\tilde{\varepsilon}_t(y, x; \theta) = \log D^{\mathcal{E}}_t(x; \theta) - p^M_t + p^Y_t - m + y$$

By the properties of probability densities, the density of y_{jt} conditional on x^*_{jt} and $x^*_{j,t-1}$ is

$$f_{y|x^*,x^*_{-1}}(y_{jt}|x^*_{jt}, x^*_{j,t-1}) = f_{\varepsilon|x^*,x^*_{-1}}(\tilde{\varepsilon}_t(y_{jt}, x^*_{jt}; \theta^0)|x^*_{jt}, x^*_{j,t-1}) \left| \frac{\partial}{\partial y} \tilde{\varepsilon}_t(y_{jt}, x^*_{jt}; \theta^0) \right|$$

with $\partial\tilde{\varepsilon}_t(y, x; \theta^0)/\partial y = 1$. This result is intuitive: given current period inputs, I can invert for persistent productivity ω, and given inputs and ω, the only variation left in y is ε

The density of y conditional on current and lagged inputs can therefore be represented in terms of the production function f and the density of the transient shock f_ε. It is possible to estimate f_ε and $\partial f/\partial m$ by maximum likelihood using this density alone, analogous to the way the first stage of GNR, alone, recovers $\partial f/\partial m$ and $\hat{\varepsilon}_{jt}$. In fact, when ε is assumed normal and f takes any parametric functional form, setting the gradient of $\sum_{j=1}^{N}\sum_{t=0}^{T-1}\log f_{\varepsilon,t}(y_{jt}, x^*_{jt};\theta)$ with respect to the parameters of f equal to zero yields an expression nearly identical to the sample analogue of the first-stage moment condition of GNR, except that, as noted in Section 3.2, $\mathcal{E}$ is computed from f_ε rather than as a sample average.

Next, the density of intermediates given current capital and labor and all lagged inputs can be written in terms of the density of the productivity innovation η to construct a maximum-likelihood analogue of the second stage of GNR. Recall from (11) and (12) that at the true value θ^0, $\tilde{\omega}_t(x^*_{jt};\theta^0) = \omega_{jt}$ and $\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1};\theta^0) = \eta_{jt}$. Let $f_{\eta|z^*}$ be the density of η_{jt} conditional on z^*_{jt}. Consider the change of variables from η_{jt} conditional on z^*_{jt} to m_{jt} conditional on z^*_{jt}:

$$\begin{aligned} f_{m|z^*}(m_{jt}|z^*_{jt}) &= f_{\eta|z^*}\left(\tilde{\eta}_t\left(x^*_{jt}, x^*_{j,t-1};\theta^0\right)|z^*_{jt}\right)\left|\frac{\partial}{\partial m}\tilde{\eta}_t(x^*_{jt}, x^*_{j,t-1};\theta^0)\right| \\ &= f_{\eta|z^*}\left(\tilde{\eta}_t\left(x^*_{jt}, x^*_{j,t-1};\theta^0\right)|z^*_{jt}\right)\left|\frac{\partial}{\partial m}\mathcal{Y}_t(x^*_{jt};\theta^0)\right| \end{aligned}$$

Of course, by Assumption 2, $\eta_{jt} \perp z^*_{jt}$, so

$$f_{m|z^*}(m_{jt}|z^*_{jt}) = f_\eta\left(\tilde{\eta}_t\left(x^*_{jt}, x^*_{j,t-1};\theta^0\right)\right)\left|\frac{\partial}{\partial m}\mathcal{Y}_t(x^*_{jt};\theta^0)\right|$$

so that, by Assumption 2,

$$f_{m|z^*}(m_{jt}|z^*_{jt}) = f_{\eta,t}\left(x^*_{jt}, x^*_{j,t-1};\theta^0\right)\left|\frac{\partial}{\partial m}\mathcal{Y}_t(x^*_{jt};\theta^0)\right|$$

where $f_{\eta,t}$ is as defined in the statement of Lemma 1. This change of variables assumes the possibility of inverting $\tilde{\eta}_t$ for m_t. Observe by Equation 9 that if $\mathcal{Y}_t$ is invertible in m for all k, ℓ, this inverse is well defined. As noted in GNR, under Assumption 1-3 (equivalent to their Assumptions 1, 2, and 4), there exists an intermediate demand function

$$m_{jt} = \mathbb{M}_t(k^*_{jt}, \ell_{jt}, \omega_{jt}) \tag{38}$$

which is invertible for ω_{jt}; the function is time-specific because it depends on prices, assumed to be common across firms. Invertibility of $\mathcal{Y}_t(\cdot;\theta^0)$ then follows[22] from the existence of $\mathbb{M}$ since

$$\mathbb{M}_t^{-1}(x^*_{jt}) = \omega_{jt} = \mathcal{Y}_t(x^*_{jt};\theta^0) + \mathcal{C}(k^*_{jt}, \ell_{jt};\theta^0)$$

[22]I have only established the invertibility of $\mathcal{Y}_t$ at the true $\theta^0 = (f^0, h^0, f^0_\varepsilon, f^0_\eta)$, but that is sufficient for identification. If $\mathcal{Y}$ were uninvertible at some (incorrect) $\theta \neq \theta^0$ for some observed x^* then the likelihood $f_{m|z^*}$ would be underestimated at θ for that observation, because only one value of $\mathcal{Y}$ would contribute to the calculated likelihood of m. Since the likelihood of any observation at the true θ^0 cannot be penalized this way, the possible uninvertibility of $\mathcal{Y}$ at some θ and some x^* cannot possibly cause an incorrect θ to have a higher likelihood than the true θ^0.

Consequently, the conditional likelihood of an observation $(y_{jt}, m_{jt}, z^*_{jt})$ is

$$f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt}; \theta^0) = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0) f_{\eta,t}\left(x^*_{jt}, x^*_{j,t-1}; \theta^0\right) \left| \frac{\partial}{\partial m_{jt}} \mathcal{Y}_t(x^*_{jt}; \theta^0) \right| \tag{39}$$

which is the true density of (y_{jt}, m_{jt}) given z^*_{jt} when evaluated at θ^0.

B.2 Consistency of MLE

Chen (2007) provides sufficient conditions for the identification and consistency of a sieve-nonparametric estimator. In this section I state these conditions as assumptions under which my MLE estimator satisfies those conditions. For simplicity I restrict myself to the two-period case. Define Θ as a normed function space containing the true θ^0. Since MLE recovers four functions $\theta = (f, h, f_\varepsilon, f_\eta)$, Θ is a Cartesian product of four function spaces equipped with an overall distance metric $d(\cdot, \cdot)$. A convenient choice of distance metric considers the square of the total variation distance over $f_{y,m|z^*}$ averaged over z^*:

$$d(\theta, \theta^0) = \mathrm{E}\left[d_{TV}\left(f_{y,m|z^*}(y, m|z^*; \theta), f_{y,m|z^*}(y, m|z^*; \theta^0) \right)^2 \right] \tag{40}$$

where $d_{TV}(\cdot, \cdot)$ is the total variation distance. Assume that for any $\theta \in \Theta$,f_ε, f_η are continuous densities over $\mathbb{R}$.

The true likelihood of a $\theta \in \Theta$ is

$$Q(\theta) = \mathrm{E}\left[\log f_{\varepsilon,1}(y_{j1}, x^*_{j1}; \theta) + \log f_{\eta,1}(y_{j1}, z^*_{j1}; \theta) \right]$$

Because this cannot be observed directly, define the empirical likelihood function $\hat{Q}_n : \Theta \to \mathbb{R}$ as

$$\hat{Q}_n(\theta) = \frac{1}{n} \sum_{j=1}^{n} \left(\log f_{\varepsilon,1}(y_{j1}, x^*_{j1}; \theta) + \log f_{\eta,1}(y_{j1}, z^*_{j1}; \theta) \right)$$

Because θ is infinite-dimensional, maximizing $\hat{Q}_n$ over $\theta \in \Theta$ is infeasible. Instead, define an infinite sequence of sieve spaces $\Theta_1 \subseteq \Theta_2 \subseteq \ldots \subseteq \Theta$. Then, for a sample size n, compute the estimator $\hat{\theta}_n$ by maximizing $\hat{Q}_n$ over $\Theta_{k(n)}$, where $k(n) : \mathbb{N} \to \mathbb{N}$ is a nondecreasing function mapping sample sizes to sieve spaces:

$$\hat{\theta}_n = \arg \sup_{\theta \in \Theta_{k(n)}} \hat{Q}_n(\theta)$$

Assumption 11 (Identification)**.**

1. *$Q(\theta^0) > -\infty$ and if $Q(\theta^0) = \infty$ then $Q(\theta) < \infty$ for all $\theta \in \Theta_k \setminus \{\theta^0\}$ for all $k \geq 1$.*
2. *There are nonincreasing positive functions $\delta(\cdot)$ and $g(\cdot)$ such that for all $\varepsilon > 0$ and all $k \geq 1$,*

$$Q(\theta^0) - \sup_{\theta \in \Theta_k, d(\theta, \theta^0) \geq \varepsilon} Q(\theta) \geq \delta(k) g(\varepsilon) > 0$$

Assumption 12 (Sieve spaces). $\Theta_k \subseteq \Theta_{k+1} \subseteq \Theta$ *for all* $k \geq 1$ *and there exists a sequence* $\pi_k\theta^0 \in \Theta_k$ *such that* $d(\theta^0, \pi_k\theta^0) \to 0$ *as* $k \to \infty$.

Assumption 13 (Continuity).

1. *For each* $k \geq 1$, $Q(\theta)$ *is upper semicontinuous on* Θ_k *under the metric* d
2. $|Q(\theta^0) - Q(\pi_{k(n)}\theta^0)| = o(\delta(k(n)))$

Assumption 14 (Compact sieve space). *The sieve spaces* Θ_k *are compact under* d.

Assumption 15 (Uniform convergence over series).

1. *For all* $k \gg 1$, $\text{plim}_{n\to\infty} \sup_{\theta\in\Theta_k} |\hat{Q}_n(\theta) - Q(\theta)| = 0$
2. $\hat{c}(k(n)) = o_p(\delta(k(n)))$ *where* $\hat{c}(k(n)) \equiv \sup_{\theta\in\Theta_k} |\hat{Q}_n(\theta) - Q(\theta)|$
3. $\eta_{k(n)} = o(\delta(k(n)))$

Assumption 11 concerns the nonparametric identification of θ^0. I will begin by showing how Assumption 11 is satisfied and then discuss the remaining conditions.

θ^0 is a maximizer of $Q(\theta)$ because for any z^* the Kullback-Leibler divergence

$$D_{KL}(f_{y,m|z^*}(\cdot|z^*;\theta^0), f_{y,m|z^*}(\cdot|z^*;\theta)) = Q(\theta^0) - Q(\theta) \geq 0$$

is nonnegative by Gibbs' inequality. It is also a unique maximizer. I assume that the functions in θ are each uniquely parameterized; for instance, if $f_{\varepsilon,t}(\cdot,\theta) = f_{\varepsilon,t}(\cdot,\theta^0)$ almost everywhere then $\theta = \theta^0$. Then by Lemma 3, below, if $Q(\theta^0) = Q(\theta)$, $\theta^0 = \theta$.

The very general setting of Chen (2007) accommodates cases referred to as ill-posed problems, in which there are sequences $\{\theta_k\}_{k=1}^n$ in Θ such that $\theta_k \in \Theta_k$ and $Q(\theta_k) \to Q(\theta)$ but $d(\theta_k, \theta^0)$ does not converge to zero. Under the distance metric proposed in Equation (40), estimation of θ is well-posed. Consider some $\epsilon > 0$ and θ. By Pinsker's inequality, for any z^*,

$$d_{TV}(f_{y,m|z^*}(y,m|z^*,\theta^0), f_{y,m|z^*}(y,m|z^*,\theta)) \leq \sqrt{\frac{1}{2} D_{KL}(f_{y,m|z^*}(y,m|z^*,\theta^0) \| f_{y,m|z^*}(y,m|z^*,\theta))}$$

$$d_{TV}(f_{y,m|z^*}(y,m|z^*,\theta^0), f_{y,m|z^*}(y,m|z^*,\theta))^2 \leq \frac{1}{2}\left(Q(\theta^0) - Q(\theta)\right)$$

Then, averaging over z^*,

$$\mathrm{E}[d_{TV}(f_{y,m|z^*}(y,m|z^*,\theta^0), f_{y,m|z^*}(y,m|z^*,\theta))^2] \leq \frac{1}{2}\left(Q(\theta^0) - Q(\theta)\right)$$

The left hand side is just the proposed distance metric, while the right-hand side is a constant in z^*, so

$$d(\theta,\theta^0) \leq \frac{1}{2}|Q(\theta^0) - Q(\theta)|$$

and when $\theta \neq \theta^0$, $Q(\theta^0) - Q(\theta) > 0$, so

$$d(\theta,\theta^0) < |Q(\theta^0) - Q(\theta)|$$

Therefore, Q is continuous at θ^0 and if $Q(\theta) \to Q(\theta^0)$, $\theta \to \theta^0$ by the Squeeze Theorem. This satisfies Assumption 11 with $g(\epsilon) = \epsilon$ and $\delta(k) = 1$.

An important example of an ill-posed sieve estimation problem is the nonparametric instrumental variables (NPIV) estimator of W. Newey and Powell (2003), which estimates a nonparametric function h using moment conditions of the form $\mathrm{E}[Y_1 - h(Y_2)|X] = 0$. While, at first glance, Equation (43) resembles the moment condition used by W. Newey and Powell, my nonparametric $h(\cdot)$ is a function of conditioning variables, so it is straightforward to estimate nonparametrically given conditioning variables.

Assumptions 12 and 14 are regularity conditions on the sieve spaces and are not specific to the estimator used. Assumption 12 states that the sieves are nested and it is possible to construct a sequence converging to θ^0 by choosing a point from each sieve space. Assumption 14 requires that the sieve spaces (but not necessarily the parameter space Θ) be compact. Condition 13(1) requires that Q be upper semicontinuous on Θ_k for all $k \geq 1$ under the metric d. This is not automatically satisfied by the continuity of Q at θ^0.

By Remark 3.1 of Chen (2007), when $\liminf_k \delta(k) > 0$, as it does in my case, and Q is continuous in Θ at θ^0, as I have just shown, Assumption 13(2) is satisfied by Assumption 13(1), Assumption 15(iii) is satisfied automatically and Assumption 15(ii) is implied by Assumption 15(i), which says that for any given sieve space Θ_k with large k, the maximum of the empirical likelihood $\hat{Q}_n$ over Θ_k converges to the maximum of the true likelihood Q over Θ_k.

Alternatively, by Chen's Remark 3.3, because Q is an expected value of a likelihood and $\liminf_k \delta(k) > 0$, Assumptions 13 and 15 can be replaced by the following:

Assumption 16 (Uniform convergence over series for sieve M-estimators)**.**

1. *$\{Z_j\}_{j=1}^n$ is i.i.d., $E[\sup_{\theta \in \Theta_n} |\ell(\theta, Z_j)|]$ is bounded;*
2. *There are a finite $s > 0$ and random variable $U(Z_j)$ with $E[U(Z_j)] < \infty$ such that*
$$\sup_{\theta, \theta' \in \Theta_n : d(\theta, \theta') \leq \delta} |l(\theta, Z_j) - l(\theta', Z_j)| \leq \delta^s U(Z_j)$$
3. *$\log N(\delta^{1/s}, \Theta_n, d) = o(n)$ for all $\delta > 0$.*

Lemma 3.
Suppose there exist $\theta = (f, h, f_\varepsilon, f_\eta)'$ such that

$$f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt}; \theta) = f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt}; \theta^0) \tag{41}$$

for all possible z^*_{jt}. Since the distribution of s depends only on the distributions of y and m, it follows that

$$\mathrm{E}_\theta[s_{jt}|z^*_{jt}, m_{jt}] = \mathrm{E}_{\theta^0}[s_{jt}|z^*_{jt}, m_{jt}]$$

By the law of iterated expectations

$$\mathrm{E}_\theta[s_{jt}|x^*_{jt}] = \mathrm{E}_{\theta^0}[s_{jt}|x^*_{jt}] \tag{42}$$

By Equation 5,

$$\mathrm{E}_{\theta^0}\left[s_{jt}|x^*_{jt}\right] = \log\left(\frac{\partial f^0}{\partial m}(x^*_{jt})\,\mathrm{E}_{\theta^0}[\exp(\varepsilon_{jt})]\right) + \mathrm{E}_{\theta^0}\left[\varepsilon_{jt}|x^*_{jt}\right]$$

Assumption 2 states that ε_{jt} is independent of the same-period inputs x^*_{jt}, so by the normalization that ε is mean zero, $\mathrm{E}\left[\varepsilon_{jt}|x^*_{jt}\right] = 0$, so

$$\mathrm{E}_{\theta^0}\left[s_{jt}|x^*_{jt}\right] = \log\left(\frac{\partial f^0}{\partial m}(x^*_{jt})\,\mathrm{E}_{\theta^0}[\exp(\varepsilon_{jt})]\right)$$

and f^0_ε is identified by

$$\varepsilon_{jt}(\theta^0) = s_{jt} - \mathrm{E}_0\left[s_{jt}|x^*_{jt}\right]$$

Likewise, given θ, $\varepsilon_{jt}(\theta) = s_{jt} - \mathrm{E}_\theta[s_{jt}|x^*_{jt}]$ can be recovered from the conditional expectation of s_{jt}. Then by (42), it follows that

$$\varepsilon_{jt}(\theta^0) = \varepsilon_{jt}(\theta)$$

so $f_\varepsilon = f^0_\varepsilon$ and $f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta) = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0)$. Moreover, given θ, $\mathrm{E}_\theta[\exp(\varepsilon_{jt})]$ is also known, meaning that

$$\frac{\partial f^0}{\partial m}(x^*_{jt}) = \frac{\exp\left(\mathrm{E}_{\theta^0}\left[s_{jt}|x^*_{jt}\right]\right)}{\mathrm{E}_{\theta^0}[\exp(\varepsilon_{jt})]} = \frac{\exp\left(\mathrm{E}_{\theta}\left[s_{jt}|x^*_{jt}\right]\right)}{\mathrm{E}_{\theta}[\exp(\varepsilon_{jt})]} = \frac{\partial f}{\partial m}(x^*_{jt})$$

so that the partial derivative of f with respect to m is also identified. Recall that by (9) and (6), $\mathcal{Y}$ depends only on $\partial f/\partial m$, f_ε and x^*_{jt}, so

$$\mathcal{Y}_t(x^*_{jt}; \theta) = \mathcal{Y}_t(x^*_{jt}; \theta^0)$$

Taking the conditional expectation

$$\iint \mathcal{Y}_t(x^*_{jt}; \theta^0) f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt}; \theta) dm_{jt} dy_{jt} = \iint \mathcal{Y}_t(x^*_{jt}; \theta^0) f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt}; \theta^0) dm_{jt} dy_{jt}$$

$$\mathrm{E}_\theta\left[\mathcal{Y}_t(x^*_{jt}; \theta^0)|z^*_{jt}\right] = \mathrm{E}_{\theta^0}\left[\mathcal{Y}_t(x^*_{jt}; \theta^0)|z^*_{jt}\right]$$
$$\mathrm{E}_\theta\left[\omega_{jt} - \mathcal{C}(k^*_{jt}, \ell_{jt}; \theta^0)|z^*_{jt}\right] = \mathrm{E}_{\theta^0}\left[\omega_{jt} - \mathcal{C}(k^*_{jt}, \ell_{jt}; \theta^0)|z^*_{jt}\right]$$

Then by Theorem 3 of GNR,[23] f is identified (up to an additive constant), so $f = f^0$. Finally, decompose equation 41 into

$$f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta) f_{\eta,t}(x^*_{j,t-1}, x^*_{jt}; \theta)\left|\frac{\partial}{\partial m}\mathcal{Y}_t(x^*_{jt}; \theta)\right| = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0) f_{\eta,t}(x^*_{j,t-1}, x^*_{jt}; \theta^0)\left|\frac{\partial}{\partial m}\mathcal{Y}_t(x^*_{jt}; \theta^0)\right|$$

Because $f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta) = f_{\varepsilon,t}(y_{jt}, x^*_{jt}; \theta^0)$ and $\mathcal{Y}_t(x^*_{jt}; \theta) = \mathcal{Y}_t(x^*_{jt}; \theta^0)$, whenever $f_{y,m|z^*} > 0$

$$f_{\eta,t}(z^*_{jt}; \theta) = f_{\eta,t}(z^*_{jt}; \theta^0)$$

But given f^0, f^0_ε, x^*_{jt} and $x^*_{j,t-1}$, ω_{jt} and $\omega_{j,t-1}$ can be computed, so once again taking conditional expectations

$$\mathrm{E}_{\theta^0}[\omega_{jt} - h(\omega_{j,t-1})|\omega_{j,t-1}] = \mathrm{E}_\theta[\omega_{jt} - h^0(\omega_{j,t-1})|\omega_{j,t-1}] = 0 \tag{43}$$

[23] As stated, GNR's Theorem 3 assumes that f is strictly concave and therefore nonlinear in m. However, this assumption is not used in the proof, so the important case where f is linear is not ruled out.

$$h(\omega_{j,t-1}) = h^0(\omega_{j,t-1})$$

and so $h = h^0$*. Finally,*

$$f_{\eta,t}(z^*_{jt};\theta) = f_{\eta,t}(z^*_{jt};\theta^0)$$

$$f_{\eta,t}\left(\eta_t\left(x^*_{j,t-1}, x^*_{jt}; f^0, f^0_\varepsilon, h^0\right); f_\eta\right) = f_{\eta,t}(\eta_t\left(x^*_{j,t-1}, x^*_{jt}; f^0, f^0_\varepsilon, h^0\right); f^0_\eta)$$

so $f_\eta = f^0_\eta$*. It follows that if* $f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt};\theta) = f_{y,m|z^*}(y_{jt}, m_{jt}|z^*_{jt};\theta^0)$ *almost everywhere,* $\theta^0 = \theta$*.*

B.3 Proof of Theorem 1

This statement follows from Theorem 1 of Hu and Schennach (2008). Begin by defining $y = \mathbf{u}_{j2}$, $x = k_0$, $x^* = k^*_0$, $z = m_0$, and $w = \mathbf{w_j}$. Observe that my Assumptions 5, 6, 7 correspond directly to their Assumptions 1, 3, 4. My Assumptions 8(ii)-(iii) imply that $\mathrm{E}[\xi_0|k^*_{j0}, \mathbf{w_j}] = 0$ and therefore $\mathrm{E}[k_{j0}|k^*_{j0}, \mathbf{w_j}] = k^*_{j0}$, satisfying their Assumption 5. Finally, by Lemma 2, my Assumptions 1, 2, 3, 4 and 8 imply their Assumption 2(i), while their Assumption 2(ii) is a weaker version of my Assumption 8(ii). Finally, my Equation 29 corresponds to their Equation 5. Rewriting their Equation 5 in my notation and conditioning on $\mathbf{w_j}$,

$$f_{u,k_0|m_0}(\mathbf{u}_{j2}, k_{j0}|m_{j0}, \mathbf{w_j}) \equiv \int f_{u_2|x^*_1}(\mathbf{u}_{j2}|x^*_{j1}) f_{k_0|k^*_0}(k_{j0}|k^*_{j0}, \mathbf{w_j}) f_{k^*_0|m_0}(k^*_{j0}|m_{j0}, \mathbf{w_j}) dk^*_0$$

Replacing $f_{k|k^*} = f_\xi$,

$$f_{u,k_0|m_0}(\mathbf{u}_{j2}, k_{j0}|m_{j0}, \mathbf{w_j}) \equiv \int f_{u_2|x^*_1}(\mathbf{u}_{j2}|x^*_{j1}) f_\xi(\xi_{j0}) f_{k^*_0|m_0}(k^*_{j0}|m_{j0}, \mathbf{w_j}) dk^*_0 \tag{44}$$

By Bayes' rule,

$$\begin{aligned} f_{k^*_0|m_0}(k^*_0|I_{j0}, \ell_{j0}, \ell_{j1}, m_0, m_{j1}) &= \frac{f_{k^*_{j0}, w|\ell_0, m_0}(k^*_0, I_{j0}, \ell_{j1}, m_{j1}|\ell_{j0}, m_{j0})}{f_{w|\ell_0,m_0}(I_{j0}, \ell_{j1}, m_{j1}|\ell_{j0}, m_{j0})} \\ &= \frac{f_{m|x^*}(m_{j1}|x^*_{j1}) f_{IL}(I_{j0}, \ell_{j1}|x^*_{j0}) f_{k^*|m,\ell}(k^*_{j0}|\ell_{j0}, m_0)}{f_{m_0,w}(I_{j0}, \ell_{j1}, m_{j1}|\ell_{j0}, m_{j0})} \end{aligned}$$

Substituting this into Equation 44 and rearranging

$$f_{u,k_0|m_0}(\mathbf{u}_{j2}, k_{j0}|m_{j0}, \mathbf{w_j}) = \frac{\int f_{u_1|x^*_0}(\mathbf{u}_{j1}|x^*_{j0}) f_\xi(\xi_{j0}) f_{k^*|m,\ell}(k^*_{j0}|\ell_{j0}, m_0) dk^*_0}{f_{m_0,w}(I_{j0}, \ell_{j1}, m_{j1}|\ell_{j0}, m_{j0})}$$

$$f_{u,k_0|m_0}(\mathbf{u}_{j2}, k_{j0}|m_{j0}, \mathbf{w_j}) \propto \int f_{u_1|x^*_0}(\mathbf{u}_{j2}|x^*_{j1}) f_\xi(\xi_{j0}) f_{k^*|m,\ell}(k^*_{j0}|\ell_{j0}, m_0) dk^*_0 = f_{u,w,k_0|\ell_0,m_0}$$

C Additional Results

In this Appendix I present tables of additional results, including further Monte Carlo estimates and maximum likelihood estimates on the empirical data.

C.1 Supplementary Monte Carlo Estimates

Table 5: Monte Carlo Estimates: Alternative Shock Distributions

		Baseline Model				Laplace ε				Laplace η			
	Truth	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS
Elasticities													
Capital	0.20	0.11	0.20	0.13	0.20	0.11	0.20	0.13	0.20	0.04	0.21	0.10	0.20
		(0.01)	(0.02)	(0.05)	(0.01)	(0.03)	(0.01)	(0.04)	(0.01)	(0.02)	(0.04)	(0.01)	(0.02)
Labor	0.30	0.33	0.30	0.33	0.30	0.33	0.30	0.32	0.30	0.43	0.29	0.36	0.30
		(0.01)	(0.01)	(0.06)	(0.01)	(0.02)	(0.01)	(0.05)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)
RTS	1.00	0.93	1.01	0.96	1.00	0.93	1.00	0.95	1.00	0.97	1.00	0.96	1.00
		(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Productivity													
S. D.	0.61	0.69	0.60	0.74	0.61	0.71	0.60	0.75	0.61	0.64	0.61	0.69	0.61
		(0.02)	(0.02)	(0.09)	(0.01)	(0.26)	(0.02)	(0.07)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Autocorr.	0.37	0.50	0.34	0.53	0.37	0.50	0.34	0.54	0.36	0.40	0.36	0.50	0.37
		(0.03)	(0.04)	(0.20)	(0.02)	(0.04)	(0.04)	(0.17)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)

Comparison of Monte Carlo estimates of productivity and elasticity distributions to true values under three different data-generating processes. All panels were drawn as described in Appendix A. The first panel follows the baseline DGP, for comparison. The second and third panels were drawn assuming a Laplace distribution for shock ε and innovation η, respectively. The first column in each panel gives the average standard deviation of true productivity across simulations and the true values of elasticity distributions; I simulate data assuming a Cobb-Douglas production function, so that the true elasticity is constant across firms. Four estimators are applied: error uncorrected GMM and MLE, and error-corrected GMM-HS and MLE-HS. In each case I estimate a translog production function and a degree 3 productivity process. The mean of each elasticity and the standard deviation of productivity are calculated for each estimator; I report the average of these statistics across all 200 simulations, with Monte Carlo standard deviations in parentheses.

Table 6: Monte Carlo Estimates: Heterogeneous Depreciation

		Baseline Model				Random Depreciation (i.i.d.)				Random Depreciation (fixed effect)			
	Truth	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS
Elasticities													
Capital	0.20	0.11	0.20	0.13	0.20	0.11	0.20	0.14	0.20	0.13	0.21	0.13	0.21
		(0.01)	(0.02)	(0.05)	(0.01)	(0.04)	(0.02)	(0.04)	(0.01)	(0.02)	(0.02)	(0.06)	(0.01)
Labor	0.30	0.33	0.30	0.33	0.30	0.32	0.30	0.32	0.30	0.32	0.30	0.34	0.30
		(0.01)	(0.01)	(0.06)	(0.01)	(0.01)	(0.02)	(0.05)	(0.01)	(0.01)	(0.02)	(0.07)	(0.01)
RTS	1.00	0.93	1.01	0.96	1.00	0.93	1.01	0.95	1.00	0.95	1.01	0.97	1.01
		(0.01)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)
Productivity													
S. D.	0.61	0.69	0.60	0.74	0.61	0.74	0.60	0.75	0.61	0.73	0.60	0.74	0.61
		(0.02)	(0.02)	(0.09)	(0.01)	(0.69)	(0.02)	(0.07)	(0.01)	(0.03)	(0.02)	(0.10)	(0.01)
Autocorr.	0.37	0.50	0.34	0.53	0.37	0.50	0.34	0.56	0.37	0.54	0.34	0.50	0.37
		(0.03)	(0.04)	(0.20)	(0.02)	(0.05)	(0.05)	(0.15)	(0.02)	(0.04)	(0.06)	(0.23)	(0.03)

Comparison of Monte Carlo estimates of productivity and elasticity distributions to true values under the baseline data-generating process and two alternative DGPs in which depreciation is heterogeneous across firms. All panels were drawn as described in Appendix A. In the second panel each firm draws a different i.i.d. depreciation rate in each period. The third panel was drawn by letting each firm draw a single fixed depreciation rate. The first column in each panel gives the average standard deviation of true productivity across simulations and the true values of elasticity distributions; I simulate data assuming a Cobb-Douglas production function, so that the true elasticity is constant across firms. Four estimators are applied: error uncorrected GMM and MLE, and error-corrected GMM-HS and MLE-HS. In each case I estimate a translog production function and a degree 3 productivity process. The mean of each elasticity and the standard deviation of productivity are calculated for each estimator; I report the average of these statistics across all 200 simulations, with Monte Carlo standard deviations in parentheses.

Table 7: Monte Carlo Estimates: Other Measurement Error Distributions

		Baseline Model				Wrong Depreciation				Investment Error			
	Truth	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS	GMM	GMM-HS	MLE	MLE-HS
Elasticities													
Capital	0.20	0.11	0.20	0.13	0.20	0.12	0.20	0.14	0.19	0.10	0.15	0.07	0.18
		(0.01)	(0.02)	(0.05)	(0.01)	(0.07)	(0.02)	(0.03)	(0.01)	(0.01)	(0.02)	(0.06)	(0.01)
Labor	0.30	0.33	0.30	0.33	0.30	0.32	0.30	0.31	0.30	0.33	0.31	0.39	0.30
		(0.01)	(0.01)	(0.06)	(0.01)	(0.03)	(0.02)	(0.04)	(0.01)	(0.01)	(0.01)	(0.09)	(0.01)
RTS	1.00	0.93	1.01	0.96	1.00	0.94	1.00	0.95	0.99	0.93	0.97	0.96	0.98
		(0.01)	(0.01)	(0.02)	(0.01)	(0.05)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.04)	(0.01)
Productivity													
S. D.	0.61	0.69	0.60	0.74	0.61	0.72	0.61	0.75	0.62	0.69	0.64	0.64	0.67
		(0.02)	(0.02)	(0.09)	(0.01)	(0.24)	(0.02)	(0.06)	(0.01)	(0.02)	(0.02)	(0.10)	(0.02)
Autocorr.	0.37	0.50	0.34	0.53	0.37	0.51	0.36	0.56	0.38	0.49	0.42	0.32	0.47
		(0.03)	(0.04)	(0.20)	(0.02)	(0.05)	(0.05)	(0.13)	(0.03)	(0.03)	(0.04)	(0.27)	(0.03)

Comparison of Monte Carlo estimates of productivity and elasticity distributions to true values under the baseline data-generating process and two different alternative DGPs. The alternative panels were drawn as described in Appendix A. In the second panel, the econometrician assumes an incorrect depreciation rate. In the third panel, noise was applied to the econometrician's observation of investment in each period. The first column in each panel gives the average standard deviation of true productivity across simulations and the true values of elasticity distributions; I simulate data assuming a Cobb-Douglas production function, so that the true elasticity is constant across firms. Four estimators are applied: error uncorrected GMM and MLE, and error-corrected GMM-HS and MLE-HS. In each case I estimate a translog production function and a degree 3 productivity process. The mean of each elasticity and the standard deviation of productivity are calculated for each estimator; I report the average of these statistics across all 200 simulations, with Monte Carlo standard deviations in parentheses.

C.2 Maximum Likelihood Empirical Estimates

Table 8: Maximum Likelihood Empirical Estimates

		Industry									
		311		321		322		331		381	
		MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS
Elasticities											
Capital	Mean	0.11	0.13	0.13	0.22	0.10	0.20	0.10	0.14	0.14	0.23
	S.D.	0.06	0.06	0.02	0.06	0.02	0.05	0.03	0.03	0.03	0.06
Labor	Mean	0.21	0.20	0.38	0.27	0.40	0.29	0.37	0.34	0.44	0.31
	S.D.	0.04	0.04	0.06	0.13	0.10	0.07	0.05	0.06	0.08	0.06
Interm.	Mean	0.63	0.63	0.50	0.52	0.51	0.52	0.54	0.54	0.47	0.48
	S.D.	0.03	0.03	0.05	0.10	0.08	0.10	0.04	0.04	0.07	0.10
Total	Mean	0.96	0.96	1.02	1.01	1.00	1.01	1.02	1.02	1.05	1.02
	S.D.	0.05	0.06	0.03	0.04	0.03	0.02	0.04	0.05	0.04	0.02
Capital											
	Var	4.29	4.18	3.29	2.88	2.20	1.97	3.39	2.84	3.29	3.02
Productivity											
	Var	0.11	0.11	0.14	0.12	0.12	0.11	0.16	0.16	0.16	0.14
% Bias			3.22		11.44		7.70		0.38		13.69
Firms		2739		967		972		957		1241	
Periods		18		18		18		18		18	
Firm-Period Pairs		20849		5879		5143		5432		6262	

Empirical estimates for the largest 3-digit industries in the sample under the main specification, estimated via maximum likelihood. Rows labelled "% Bias" give the percentage difference in estimated productivity variance between the error-corrected and error-uncorrected estimators.

Table 9: Maximum Likelihood Empirical Estimates: Cobb-Douglas

	Industry									
	311		321		322		331		381	
	MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS	MLE	MLE-HS
Elasticities										
Capital	0.13	0.16	0.14	0.17	0.14	0.18	0.12	0.14	0.17	0.20
Labor	0.17	0.16	0.31	0.30	0.26	0.25	0.32	0.31	0.34	0.33
Interm.	0.64	0.64	0.50	0.50	0.50	0.50	0.54	0.54	0.46	0.46
Total	0.94	0.95	0.96	0.97	0.91	0.93	0.99	0.99	0.97	0.99
Capital										
Var	4.29	3.86	3.29	2.61	2.20	1.80	3.39	2.83	3.29	2.63
Productivity										
Var	0.13	0.12	0.15	0.14	0.15	0.13	0.17	0.17	0.18	0.16
% Bias		7.87		5.32		10.60		3.56		8.78
Firms	2739		967		972		957		1241	
Periods	18		18		18		18		18	
Firm-Period Pairs	20849		5879		5143		5432		6262	

Empirical estimates for the largest 3-digit industries in the sample under a Cobb-Douglas, AR(1) specification, estimated via maximum likelihood. Rows labelled "% Bias" give the percentage difference in estimated productivity variance between the error-corrected and error-uncorrected estimators.

D Value-Added Specification

ACF consider an alternative data-generating process in which

$$y_t = \min\{\beta_K k_t^* + \beta_L \ell_t + \omega_t, \beta_M + m_t\} + \varepsilon$$

In this specification, I instead observe that

$$\varepsilon(y_t, m_t) = y_t - m_t - \beta_M$$

and

$$\tilde{\mathcal{Y}}_t = m_t + \beta_M - \beta_K k_t^* - \beta_L \ell_t$$

Observe that price drops out because of the fixed proportions imposed by the production function and that the $\left|\frac{\partial}{\partial m}\tilde{\mathcal{Y}}_t(\cdot;\theta)\right|$ term is equal to 1.

$$f_{y,m|z^*} = f_{\varepsilon,t} f_{\eta,t}$$

However, ACF's DGP1 violates my assumption 10 by allowing ℓ to be chosen with partial information about same-period η. The amount of "extra" information available to the firm in choosing labor is controlled by their parameter b. To test my estimator using this data-generating process, I modify it by setting $b = 1$, corresponding to my assumption that labor is chosen a full period in advance. Under this additional assumption, my estimator performs well on ACF's Monte Carlo data, suggesting that my approach can be applied to the value-added setting as well.

References

Ackerberg, D. A. (2023). Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context. *The Journal of Industrial Economics*, *71*(3), 644–674.

Ackerberg, D. A., Caves, K., & Frazer, G. (2015). Identification Properties of Recent Production Function Estimators. *Econometrica*, *83*(6), 2411–2451.

Aguirregabiria, V., & Mira, P. (2007). Sequential Estimation of Dynamic Discrete Games. *Econometrica*, *75*(1), 1–53.

Amiti, M., & Konings, J. (2007). Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia. *American Economic Review*, *97*(5), 1611–1638.

Aw, B. Y., Roberts, M. J., & Xu, D. Y. (2008). R&D Investments, Exporting, and the Evolution of Firm Productivity. *The American Economic Review*, *98*(2), 451–456.

Bartelsman, E. J., & Doms, M. (2000). Understanding Productivity: Lessons from Longitudinal Microdata. *Journal of Economic Literature*, *38*(3), 569–594. Retrieved November 4, 2024, from http://www.jstor.org/stable/2565420

Becker, R. A., Haltiwanger, J., Jarmin, R. S., Klimek, S. D., & Wilson, D. J. (2006). Micro and Macro Data Integration: The Case of Capital. In *A New Architecture for the U.S. National Accounts* (pp. 541–610). University of Chicago Press.

Bernard, A. B., Eaton, J., Jensen, J. B., & Kortum, S. (2003). Plants and Productivity in International Trade. *American Economic Review*, *93*(4), 1268–1290.

Biesebroeck, J. V. (2003). Productivity Dynamics with Technology Choice: An Application to Automobile Assembly. *The Review of Economic Studies*, *70*(1), 167–198.

Chen, X. (2007). Large Sample Sieve Estimation of Semi-Nonparametric Models (J. Heckman & E. Leamer, Eds.; 1st ed.). In J. Heckman & E. Leamer (Eds.), *Handbook of Econometrics* (1st ed.). Elsevier.

Chen, X., Christensen, T., & Kankanala, S. (2024). Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities. *The Review of Economic Studies*.

Collard-Wexler, A., & De Loecker, J. (2024). *Production Function Estimation with Measurement Error in Inputs* (NBER Working Papers No. 22437). National Bureau of Economic Research, Inc.

De Loecker, J. (2011). Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity. *Econometrica*, *79*, 1407–1451.

De Ridder, M., Grassi, B., & Morzenti, G. (2026). The Hitchhiker's Guide to Markup Estimation: Assessing Estimates From Financial Data. *Econometrica*, *94*(1), https://onlinelibrary.wiley.com 137–168. https://doi.org/https://doi.org/10.3982/ECTA22733

Demirer, M. (2023). Production Function Estimation with Imperfect Proxies.

Dollar, D., Hallward-Driemeier, M., & Mengistae, T. (2005). Investment Climate and Firm Performance in Developing Economies. *Economic Development and Cultural Change*, *54*(1), 1–31.

Dong, B., & Hsieh, Y.-W. (2021). Misallocation or Mismeasurement? Evidence from Plant-Level Data. *Annals of Economics and Statistics*, (142), 211–234. Retrieved November 25, 2025, from https://www.jstor.org/stable/10.15609/annaeconstat2009.142.0211

Doraszelski, U., & Jaumandreu, J. (2018). Measuring the Bias of Technological Change. *Journal of Political Economy*, *126*(3), 1027–1084. https://EconPapers.repec.org/RePEc:ucp:jpolec:doi:10.1086/697204

Galuscak, K., & Lizal, L. (2011). *The Impact of Capital Measurement Error Correction on Firm-Level Production Function Estimation* (Working Papers). Czech National Bank.

Gandhi, A., Navarro, S., & Rivers, D. A. (2020). On the Identification of Gross Output Production Functions. *Journal of Political Economy*, *128*(8), 2973–3016.

Gorodnichenko, Y. (2007). *Using Firm Optimization to Evaluate and Estimate Returns to Scale* (Working Paper No. 13666). National Bureau of Economic Research.

Greenstreet, D. (2007). *Appendices for Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules* (Economics Series Working Papers No. 345). University of Oxford, Department of Economics.

Griliches, Z., & Mairesse, J. (1995). *Production Functions: The Search for Identification* (Working Paper No. 5067). National Bureau of Economic Research.

Halpern, L., Koren, M., & Szeidl, A. (2015). Imported Inputs and Productivity. *American Economic Review*, *105*(12), 3660–3703.

Hicks, J. (1981). *Wealth and welfare*. Oxford, Blackwell.

Hotz, V. J., & Miller, R. A. (1993). Conditional Choice Probabilities and the Estimation of Dynamic Models. *The Review of Economic Studies*, *60*(3), 497–529.

Hu, Y., & Schennach, S. (2008). Instrumental Variable Treatment of Nonclassical Measurement Error Models. *Econometrica*, *76*(1), 195–216.

Kasahara, H., & Rodrigue, J. (2008). Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence. *Journal of Development Economics*, *87*(1), 106–118.

Kátay, G., & Wolf, Z. (2008). *Driving Factors of Growth in Hungary - a Decomposition Exercise* (MNB Working Papers No. 2008/6). Magyar Nemzeti Bank (Central Bank of Hungary). https://doi.org/None

Kim, K., Petrin, A., & Song, S. (2016). Estimating Production Functions with Control Functions When Capital is Measured with Error. *Journal of Econometrics*, *190*(2), 267–279.

Levinsohn, J., & Petrin, A. (2003). Estimating Production Functions Using Inputs to Control for Unobservables. *The Review of economic studies*, *70*(2), 317–341.

Liu, L. (1991). *Entry-Exit and Productivity Change: An Empirical Analysis of Efficiency Frontiers* (Doctoral dissertation). University of Michigan.

Mueller, S. (2008). A Modified Perpetual Inventory Approach. *Jahrbücher für Nationalökonomie und Statistik*, *228*(4), 357–371. https://doi.org/doi:10.1515/jbnst-2008-0405

Newey, W. K. (2013). Nonparametric Instrumental Variables Estimation. *American Economic Review*, *103*(3), 550–556.

Newey, W., & Powell, J. (2003). Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, *71*(5), 1565–1578.

OECD. (2009). *Measuring Capital - OECD Manual* (Second Edition, tech. rep.). Paris, OECD Publishing.

Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, *64*(6), 1263–1297. Retrieved May 14, 2026, from http://www.jstor.org/stable/2171831

Pavcnik, N. (2002). Trade Liberalization, Exit, and Productivity Improvements: Evidence from Chilean Plants. *The Review of Economic Studies*, *69*(1), 245–276.

Schennach, S. M. (2004). Estimation of Nonlinear Models with Measurement Error. *Econometrica*, *72*(1), https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1468-0262.2004.00477.x, 33–75. https://doi.org/https://doi.org/10.1111/j.1468-0262.2004.00477.x

Sivadasan, J. (2009). Barriers to Competition and Productivity: Evidence from India. *The B.E. Journal of Economic Analysis & Policy*, *9*(1).

Stevens, G. V. G. (1994). Measurement Error in the Capital Stock and Its Effect on the Investment Function. *Applied Economics*, *26*(1), 21–26.

Syverson, C. (2001). *Output Market Structure and Productivity Heterogeneity*. University of Maryland at College Park.

Syverson, C. (2011). What Determines Productivity? *Journal of Economic Literature*, *49*(2), 326–65. https://doi.org/10.1257/jel.49.2.326

Topalova, P., & Khandelwal, A. (2011). Trade Liberalization and Firm Productivity: The Case of India. *The Review of Economics and Statistics*, *93*(3), 995–1009.

Wooldridge, J. (2009). On Estimating Firm-level Production Functions Using Proxy Variables to Control for Unobservables. *Economics Letters*, *104*(3), 112–114.