On the Identification of Gross Output Production Functions

Amit Gandhi

University of Pennsylvania

Salvador Navarro

University of Western Ontario

David A. Rivers

University of Western Ontario

We study the nonparametric identification of gross output production functions under the environment of the commonly employed proxy variable methods. We show that applying these methods to gross output requires additional sources of variation in the demand for flexible inputs (e.g., prices). Using a transformation of the firm's first-order condition, we develop a new nonparametric identification strategy for gross output that can be employed even when additional sources of variation are not available. Monte Carlo evidence and estimates from Colombian and Chilean plant-level data show that our strategy performs well and is robust to deviations from the baseline setting.

I. Introduction

The identification and estimation of production functions using data on firm inputs and output are among the oldest empirical problems in

We thank Dan Ackerberg, Roy Allen, Richard Blundell, Juan Esteban Carranza, Allan Collard-Wexler, Ulrich Doraszelski, Steven Durlauf, Jeremy Fox, Silvia Goncalves, Phil Haile, Jinyong Hahn, Joel Horowitz, Jean-Francois Houde, Nail Kashaev, Lance Lochner,

Electronically published June 8, 2020
[*Journal of Political Economy*, 2020, vol. 128, no. 8]

economics. A key challenge for identification arises because firms optimally choose their inputs as a function of their productivity, but productivity is unobserved by the econometrician. As first articulated by Marschak and Andrews (1944), this gives rise to a simultaneity problem that is known in the production function literature as "transmission bias." Solving this identification problem is critical to measuring productivity with plant-level production data, which have become increasingly available for many countries and which motivate a variety of industry equilibrium models based on patterns of productivity heterogeneity found in these data.[1]

The recent literature on production function estimation focuses on environments in which some inputs satisfy static first-order conditions (flexible inputs) and some do not. In this paper, we study the nonparametric identification of gross output production functions in this setting. We clarify the conditions under which existing estimators can be applied. We then propose an alternative, nonparametric identification and estimation strategy that does not rely on having access to exogenous price variation or other exclusion restrictions (e.g., policy variation).

As discussed in their influential review of the state of the literature, Griliches and Mairesse (1998) concluded that the standard econometric solutions to correct the transmission bias—that is, using firm fixed effects or instrumental variables—are both theoretically problematic and unsatisfactory in practice (for a more recent review, see also Ackerberg et al. 2007). An alternative early approach to addressing the simultaneity problem employed static first-order conditions for input choices. The popular index number methods (see, e.g., Caves, Christensen, and Diewert 1982) recover the production function and productivity by equating each input's output elasticity to its input share. However, when some inputs

Aureo de Paula, Amil Petrin, Mark Roberts, Nicolas Roys, Todd Stinebrickner, Chad Syverson, Chris Taber, Quang Vuong, and especially Tim Conley and Paul Schrimpf for helpful discussions. Cecilia Diaz Campo provided outstanding research assistance. This paper has also benefited from detailed comments by the editor and three anonymous referees. We also thank Amil Petrin and David Greenstreet for helping us to obtain the Colombian and Chilean data, respectively. Navarro and Rivers acknowledge support from the Social Sciences and Humanities Research Council of Canada. This paper previously circulated under the names "Identification of Production Functions Using Restrictions from Economic Theory" and "On the Identification of Production Functions: How Heterogeneous Is Productivity?" and included what is now a companion paper titled "How Heterogeneous Is Productivity? A Comparison of Gross Output and Value Added." Data are provided as supplementary material online.

[1] Among these patterns are the general understanding that even narrowly defined industries exhibit "massive" unexplained productivity dispersion (Dhrymes 1991; Bartelsman and Doms 2000; Syverson 2004; Collard-Wexler 2010; Fox and Smeets 2011) and that productivity is closely related to other dimensions of firm-level heterogeneity, such as importing (Kasahara and Rodrigue 2008), exporting (Bernard and Jensen 1995, 1999; Bernard et al. 2003), wages (Baily, Hulten, and Campbell 1992), etc. See Syverson (2011) for a review of this literature.

are subject to adjustment frictions, such as adjustment costs for capital or hiring/firing costs for labor, these static first-order conditions are no longer valid.[2]

More recently, the literature on production function estimation has studied settings in which not all inputs satisfy static first-order conditions and thus standard index number methods cannot be applied. Instead, transmission bias is addressed by imposing assumptions on the economic environment, which allows researchers to exploit lagged input decisions as instruments for current inputs. This strategy is fundamental to both of the main strands of structural estimation approaches—namely, dynamic panel methods (Arellano and Bond 1991; Blundell and Bond 1998, 2000), as well as the proxy variable methods (Olley and Pakes 1996; Levinsohn and Petrin 2003; Wooldridge 2009; Ackerberg, Caves, and Frazer 2015) that are now prevalent in the applied literature on production function estimation.[3] Most of these papers (with the exception of Levinsohn and Petrin 2003) focus on some form of a value-added production function. Recent work by Ackerberg, Caves, and Frazer (2015) has carefully examined the identification foundations of these estimators in the context of value added. No such analysis has been done for gross output.[4]

Recently, however, there has been a growing interest in estimating gross output models of production. In the international trade literature, researchers are studying the importance of imported intermediate inputs for productivity (Amiti and Konings 2007; Kasahara and Rodrigue 2008; Halpern, Koren, and Szeidl 2015; De Loecker et al. 2016). The macroeconomics literature on misallocation is also now employing gross output models of firm-level production (Oberfield 2013; Bils, Klenow, and Ruane 2017). As another example, papers interested in separating the importance of productivity from demand-side heterogeneity (e.g., markups and demand shocks) are using gross output production functions (Foster, Haltiwanger, and Syverson 2008; Pozzi and Schivardi 2016; Blum et al. 2019). While in principle the proxy variable and dynamic panel methods can be extended to estimate gross output forms of the production function, the identification of such an approach has not been systematically examined.

We begin by studying the nonparametric identification of these structural methods extended to gross output. Our first main result is to show

[2] Alternatively, these approaches can avoid imposing this assumption for one input by imposing restrictions on returns to scale, often assuming constant returns to scale.

[3] While the term "proxy variable approach" could encompass a wide variety of methods (see, e.g., Heckman and Robb 1985), throughout this paper when we refer to proxy variable methods, we mean those of Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg, Caves, and Frazer (2015).

[4] For a discussion of the relationship between gross output and value-added production functions, see Bruno (1978), Basu and Fernald (1997), and Gandhi, Navarro, and Rivers (2017).

that, absent sources of variation in flexible input demand other than a panel of data on output and inputs, the gross output production function is nonparametrically nonidentified under these approaches. We then show that under the assumption that the model structure (e.g., the production function) does not vary over time, time-series variation in aggregate price indexes presents a potential source of identifying variation. However, Monte Carlo evidence suggests that this source of variation, while valid in theory, may perform poorly in practice, even in relatively long panels. In the context of a parametric setting, Doraszelski and Jaumandreu (2013) provide an alternative solution that instead incorporates observed firm-level variation in prices. In particular, they show that by explicitly imposing the parameter restrictions between the production function and the demand for a flexible input (which underlies the proxy variable approaches of Levinsohn and Petrin 2003 and Ackerberg, Caves, and Frazer 2015) and by using this price variation, they can recover the gross output production function.

Our second contribution is that we present a new empirical strategy that nonparametrically identifies the gross output production function. Our strategy is particularly useful for (but not limited to) settings in which researchers do not have access to long panels with rich aggregate time-series price variation or access to firm-specific prices or other external instruments. As in Doraszelski and Jaumandreu (2013), we recognize the structural link between the production function and the firm's first-order condition for a flexible input.[5] The key to our approach is that we exploit this link in a fully nonparametric setting. In particular, we show that a nonparametric regression of the flexible input's revenue share on all inputs (labor, capital, and intermediate inputs) identifies the flexible input elasticity. We then recognize that the flexible input elasticity defines a partial differential equation on the production function, which imposes nonparametric cross-equation restrictions with the production function itself. We can solve this partial differential equation to nonparametrically identify the part of the production function that depends on the flexible input. This is a nonparametric analogue of the familiar parametric insight that revenue shares directly identify the flexible input coefficient in a Cobb-Douglas setting (e.g., Klein 1953; Solow 1957). We then use the dynamic panel/proxy variable conditional moment restrictions based on lagged input decisions for the remaining inputs. By combining insights from the index number literature (using shares) with those from the dynamic panel literature (using lags as instruments), we

[5] The use of optimality conditions to exploit cross-equation restrictions for identification is well established in economics. See, e.g., Heckman (1974) for labor supply, Hansen and Singleton (1982) for consumption, and Lucas and Sargent (1981) for many examples of the use of cross-equation restrictions in the context of the rational expectations literature.

show that the gross output production function and productivity can be nonparametrically identified.

This identification strategy—regressing revenue shares on inputs to identify the flexible input elasticity, solving the partial differential equation, and integrating this into the dynamic panel/proxy variable structure to identify the remainder of the production function—gives rise to a natural two-step nonparametric sieve estimator in which different components of the production function are estimated via polynomial series in each stage. We present a computationally straightforward implementation of this estimator. Furthermore, as the numerical equivalence result in Hahn, Liao, and Ridder (2018) shows, our estimator has the additional advantage that inference on functionals of interest can be performed using standard two-step parametric results. This gives us a straightforward approach to inference.

We validate the performance of our empirical strategy on simulated data generated under three different production functions (Cobb-Douglas, constant elasticity of substitution [CES], and translog). We find that our nonparametric estimator performs quite well in all cases. We also show that our procedure is robust to misspecification arising from the presence of adjustment costs in the flexible input. We then apply our estimator, as well as several extensions of it, to plant-level data from Colombia and Chile. We show that our estimates correct for transmission bias present in ordinary least squares (OLS). Consistent with the presence of transmission bias, OLS overestimates the flexible intermediate-input elasticities and underestimates the elasticities of capital and labor. OLS estimates also tend to understate the degree of productivity heterogeneity compared with our estimates. Finally, we show that our estimates are robust to allowing for fixed effects, alternative flexible inputs, or some additional unobservables in the flexible-input demand.

The rest of the paper is organized as follows. In section II, we describe the model and set up the firm's problem. In section III, we examine the extent to which the proxy variable/dynamic panel methods can be applied to identify the gross output production function. Section IV presents our nonparametric identification strategy. In section V, we describe our estimation strategy. Section VI compares our approach with the related literature. In section VII, we present estimates from our procedure applied to Monte Carlo simulated data as well as plant-level data from Colombia and Chile. Section VIII concludes.

II. The Model

In this section, we describe the economic model of the firm that we study. We focus attention in the main text on the classic case of perfect competition in the intermediate-input and output markets. We discuss the case of monopolistic competition with unobserved output prices in appendix O6 (apps. O1–O8 are available online).

A. Data and Definitions

We observe a panel consisting of firms $j = 1, \ldots, J$ over periods $t = 1, \ldots, T$. A generic firm's output, capital, labor, and intermediate inputs are denoted by $(Y_{jt}, K_{jt}, L_{jt}, M_{jt})$, respectively, and their log values are denoted in lowercase by $(y_{jt}, k_{jt}, l_{jt}, m_{jt})$. Firms are sampled from an underlying population, and the asymptotic dimension of the data is to let the number of firms $J \to \infty$ for a fixed T; that is, the data take a short panel form. From this data, the econometrician can directly recover the joint distribution of $\{(y_{jt}, k_{jt}, l_{jt}, m_{jt})\}_{t=1}^{T}$.

Firms have access to information in period t, which we model as a set of random variables $\mathcal{I}_{jt}$.[6] The information set $\mathcal{I}_{jt}$ contains the information that the firm can use to solve its period t decision problem. Let $x_{jt} \in \{k_{jt}, l_{jt}, m_{jt}\}$ denote a generic input. If an input is such that $x_{jt} \in \mathcal{I}_{jt}$—that is, the amount of the input employed in period t is in the firm's information set for that period—then we say that the input is *predetermined* in period t. Thus, a predetermined input is a function of the information set of a prior period, $x_{jt} = \mathbb{X}(\mathcal{I}_{jt-1}) \in \mathcal{I}_{jt}$. If an input's optimal period t choices are affected by lagged values of that same input, then we say that the input is *dynamic*. If an input is neither predetermined nor dynamic, then we say that it is *flexible*. We refer to inputs that are predetermined, dynamic, or both as *nonflexible*.

B. The Production Function and Productivity

We assume that the relationship between output and inputs is determined by an underlying production function F and a Hicks neutral productivity shock ν_{jt}.

ASSUMPTION 1. The relationship between output and the inputs takes the form

$$\begin{aligned} Y_{jt} &= F(k_{jt}, l_{jt}, m_{jt})e^{\nu_{jt}} \Leftrightarrow \\ y_{jt} &= f(k_{jt}, l_{jt}, m_{jt}) + \nu_{jt}. \end{aligned} \tag{1}$$

The production function f is differentiable at all $(k, l, m) \in \mathbb{R}^3_{++}$ and strictly concave in m.

Following the proxy variable literature, the Hicks neutral productivity shock ν_{jt} is decomposed as $\nu_{jt} = \omega_{jt} + \varepsilon_{jt}$. The distinction between ω_{jt} and

[6] Formally, the firm's information set is the σ-algebra $\sigma(\mathcal{I}_{jt})$ spanned by these random variables $\mathcal{I}_{jt}$. For simplicity, we refer to $\mathcal{I}_{jt}$ as the information set.

ε_{jt} is that ω_{jt} is known to the firm before making its period t decisions, whereas ε_{jt} is an ex post productivity shock realized only after period t decisions are made. The stochastic behavior of both of these components is explained next.

Assumption 2. $\omega_{jt} \in \mathcal{I}_{jt}$ is known to the firm at the time of making its period t decisions, whereas $\varepsilon_{jt} \notin \mathcal{I}_{jt}$ is not. Furthermore, ω_{jt} is Markovian so that its distribution can be written as $P_{\omega}(\omega_{jt} \mid \mathcal{I}_{jt-1}) = P_{\omega}(\omega_{jt} \mid \omega_{jt-1})$. The function $h(\omega_{jt-1}) = E[\omega_{jt} \mid \omega_{jt-1}]$ is continuous. The shock ε_{jt}, on the other hand, is independent of the within-period variation in information sets, $P_{\varepsilon}(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_{\varepsilon}(\varepsilon_{jt})$.

Given that $\omega_{jt} \in \mathcal{I}_{jt}$ but ε_{jt} is completely unanticipated on the basis of $\mathcal{I}_{jt}$, we will refer to ω_{jt} as persistent productivity, ε_{jt} as ex post productivity, and $\nu_{jt} = \omega_{jt} + \varepsilon_{jt}$ as total productivity. Observe that we can express $\omega_{jt} = h(\omega_{jt-1}) + \eta_{jt}$, where η_{jt} satisfies $E[\eta_{jt} \mid \mathcal{I}_{jt-1}] = 0$. The random variable η_{jt} can be interpreted as the (unanticipated at period $t-1$) "innovation" to the firm's persistent productivity ω_{jt} in period t.[7]

Without loss of generality, we can normalize $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = E[\varepsilon_{jt}] = 0$, which is in units of log output. However, the expectation of the ex post shock, in units of the level of output, becomes a free parameter that we denote as $\mathcal{E} \equiv E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}] = E[e^{\varepsilon_{jt}}]$.[8] As opposed to the independence assumption on ε_{jt} in assumption 2, much of the previous literature assumes only mean independence $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$ explicitly (although stronger implicit assumptions are imposed, as we discuss below). This distinction would be important if more capital-intensive firms faced less volatile ex post productivity shocks due to automation, for example. In terms of our analysis, the only role that full independence plays (relative to mean independence) is allowing us to treat $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}]$ as a constant, which makes the analysis more transparent.[9] If only mean independence is assumed, we would have $\mathcal{E}(\mathcal{I}_{jt}) \equiv E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}]$. We discuss the implications of this distinction below in our discussion of assumption 4 for proxy variable methods and after theorem 2 for our proposed identification strategy.

Our interest is in the case in which the production function contains both flexible and nonflexible inputs. For simplicity, we mainly focus on the case of a single flexible input in the model (but see app. O6)—namely, intermediate inputs m—and treat capital k and labor l as predetermined in the model (hence, k_{jt}, $l_{jt} \in \mathcal{I}_{jt}$). We could have also generalized the

[7] It is straightforward to allow the distribution of $P_{\omega}(\omega_{jt} \mid \mathcal{I}_{jt-1})$ to depend on other elements of $\mathcal{I}_{jt-1}$, such as firm export or import status, R&D, etc. In these cases, ω_{jt} becomes a controlled Markov process from the firm's point of view. See Kasahara and Rodrigue (2008) and Doraszelski and Jaumandreu (2013) for examples.

[8] See Goldberger (1968) for an early discussion of the implicit reinterpretation of results that arises from ignoring $\mathcal{E}$ (i.e., setting $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}] = 1$ while simultaneously setting $E[\varepsilon_{jt}] = 0$) in the context of Cobb-Douglas production functions.

[9] While independence is sufficient, we could replace this assumption with mean independence and the high-level assumption that $\mathcal{E} \equiv E[e^{\varepsilon_{jt}}]$ is a constant.

model to allow it to vary with time t (e.g., f_t, h_t). For the most part, we do not use this more general form of the model in the analysis to follow because the added notational burden distracts from the main ideas of the paper. However, we revisit this idea below when it is particularly relevant for our analysis.

C. The Firm's Problem

The proxy variable literature of Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg, Caves, and Frazer (2015) uses a flexible input demand—intermediate inputs—to proxy for the unobserved persistent productivity ω.[10] To do so, they assume that the demand for intermediate inputs can be written as a function of a single unobservable (ω), the so-called *scalar unobservability* assumption,[11] and that the input demand is *strictly monotone* in ω (see, e.g., assumptions 4 and 5 in Ackerberg, Caves, and Frazer 2015). We formalize this in the following assumption.

Assumption 3. The scalar unobservability and strict monotonicity assumptions of Levinsohn and Petrin (2003), Wooldridge (2009), and Ackerberg, Caves, and Frazer (2015) place the following restriction on the flexible input demand:

$$m_{jt} = \mathbb{M}_t\left(k_{jt}, l_{jt}, \omega_{jt}\right). \tag{2}$$

The intermediate-input demand $\mathbb{M}$ is assumed to be strictly monotone in a single unobservable ω_{jt}.

We follow the same setup used by both Levinsohn and Petrin (2003) and Ackerberg, Caves, and Frazer (2015) to justify assumption 3.[12] In particular, we write down the same problem of a profit-maximizing firm under perfect competition. From this, we derive the explicit intermediate-input demand equation underlying assumption 3. The following assumption formalizes the environment in which firms operate.

Assumption 4. Firms are price takers in the output and intermediate-input markets, with ρ_t denoting the common intermediate-input price and P_t denoting the common output price facing all firms in period t. Firms maximize expected discounted profits.

Under assumptions 1, 2, and 4, the firm's profit-maximization problem with respect to intermediate inputs is

$$\max_{M_{jt}} P_t E\left[F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}+\varepsilon_{jt}} \mid \mathcal{I}_{jt}\right] - \rho_t M_{jt}, \tag{3}$$

[10] See Heckman and Robb (1985) for an early exposition (and Heckman and Vytlacil 2007 for a general discussion) of the replacement function approach of using observables to perfectly proxy for unobservables.

[11] Olley and Pakes (1996) do not include intermediate inputs in their model.

[12] See app. A in Levinsohn and Petrin (2003) and p. 2429 in Ackerberg, Caves, and Frazer (2015).

which follows because M_{jt} does not have any dynamic implications and thus affects only current-period profits. The first-order condition of problem (3) is

$$P_t \frac{\partial}{\partial M_{jt}} F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}} \mathcal{E} = \rho_t . \tag{4}$$

This equation can then be used to solve for the demand for intermediate inputs

$$m_{jt} = \mathbb{M}\left(k_{jt}, l_{jt}, \omega_{jt} - d_t\right) = \mathbb{M}_t\left(k_{jt}, l_{jt}, \omega_{jt}\right), \tag{5}$$

where $d_t \equiv \ln(\rho_t/P_t) - \ln \mathcal{E}$. It can also be inverted to solve for productivity, ω.

Equations (4) and (5) are derived under the assumption that ε_{jt} is independent of the firm's information set $(P_\varepsilon(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_\varepsilon(\varepsilon_{jt}))$. If instead only mean independence of ε_{jt} were assumed $(E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0)$, we would have $P_t(\partial F(k_{jt}, l_{jt}, m_{jt})/\partial M_{jt})e^{\omega_{jt}}\mathcal{E}(\mathcal{I}_{jt}) = \rho_t$, and hence $m_{jt} = \mathbb{M}_t(k_{jt}, l_{jt}, \omega_{jt}, \mathcal{I}_{jt})$. Assumption 3 is therefore implicitly imposing that if $\mathcal{E}(\mathcal{I}_{jt})$ is not constant, then it is at most a function of the variables already included in equation (2). In theory, this can be relaxed by allowing the proxy equation to also depend on the other elements of the firm's information set, as long as this is done in a way that does not violate scalar unobservability/monotonicity.

Given the structure of the production function, we can formally state the problem of transmission bias in the nonparametric setting. Transmission bias classically refers to the bias in Cobb-Douglas production function parameter estimates from an OLS regression of output on inputs (see Marschak and Andrews 1944; Griliches and Mairesse 1998). In the nonparametric setting, we can see transmission bias more generally as the empirical problem of regressing output y_{jt} on inputs (k_{jt}, l_{jt}, m_{jt}), which yields

$$E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = f\left(k_{jt}, l_{jt}, m_{jt}\right) + E\left[\omega_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right],$$

and hence the elasticity of the regression in the data with respect to an input $x_{jt} \in \{k_{jt}, l_{jt}, m_{jt}\}$,

$$\frac{\partial}{\partial x_{jt}} E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = \frac{\partial}{\partial x_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) + \frac{\partial}{\partial x_{jt}} E\left[\omega_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right],$$

is a biased estimate of the true production elasticity $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial x_{jt})$.

III. The Proxy Variable Framework and Gross Output

Both the dynamic panel literature and the proxy literature of Olley and Pakes (1996), Levinsohn and Petrin (2003), Wooldridge (2009), and

Ackerberg, Caves, and Frazer (2015) have mainly focused on estimating value-added models of production, in which intermediate inputs do not enter the estimated production function.[13] One exception is Levinsohn and Petrin (2003), which employs a gross output specification. However, previous work by Bond and Söderbom (2005) and Ackerberg, Caves, and Frazer (2015) has identified an identification problem with the Levinsohn and Petrin (2003) procedure. Therefore, in this section we examine whether the modified proxy variable approach developed by Ackerberg, Caves, and Frazer (2015) for value-added production functions can be extended to identify gross output production functions under the setup described in the previous section.[14]

Under the proxy variable structure, the inverted proxy equation, $\omega_{jt} = \mathbb{M}^{-1}(k_{jt}, l_{jt}, m_{jt}) + d_t$, is used to replace for productivity. Here transmission bias takes a very specific form:

$$\begin{aligned} E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, d_t\right] &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathbb{M}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t \\ &\equiv \phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t, \end{aligned} \tag{6}$$

where d_t represents a time fixed effect. Clearly, no structural elasticities can be identified from this regression (the "first stage")—in particular, the flexible input elasticity, $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$. Instead, all the information from the first stage is summarized by the identification of the random variable $\phi(k_{jt}, l_{jt}, m_{jt})$ and, as a consequence, the ex post productivity shock $\varepsilon_{jt} = y_{jt} - E[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, d_t]$.

The question then becomes whether the part of $\phi(k_{jt}, l_{jt}, m_{jt})$ that is due to $f(k_{jt}, l_{jt}, m_{jt})$ versus the part that is due to ω_{jt} can be separately identified using the second-stage restrictions of the model. This second stage is formed by adopting a key insight from the dynamic panel data literature (Arellano and Bond 1991; Blundell and Bond 1998, 2000)—namely, that given an assumed time-series process for the unobservables (in this case, the Markovian process for ω in assumption 2), appropriately lagged input decisions can be used as instruments. That is, we can rewrite the production function as

[13] Intermediate inputs, however, may still be used as the proxy variable for productivity (see Ackerberg, Caves, and Frazer 2015).

[14] We restrict our attention in the main text to the use of intermediate inputs as a proxy vs. the original proxy variable strategy of Olley and Pakes (1996) that uses investment. As Levinsohn and Petrin (2003) argued, the fact that investment is often zero in plant-level data leads to practical challenges in using the Olley and Pakes (1996) approach, and as a result, using intermediate inputs as a proxy has become the preferred strategy in applied work. In app. O1, we show that our results extend to the case of using investment instead, as well as to the use of dynamic panel methods.

$$\begin{aligned} y_{jt} &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \omega_{jt} + \varepsilon_{jt} \\ &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \quad (7) \\ &\quad + \eta_{jt} + \varepsilon_{jt} \end{aligned}$$

to form the second-stage equation. Assumption 2 implies that for any transformation $\Gamma_{jt} = \Gamma(\mathcal{I}_{jt-1})$ of the lagged-period information set $\mathcal{I}_{jt-1}$ we have the orthogonality $E[\eta_{jt} + \varepsilon_{jt} \mid \Gamma_{jt}] = 0$.[15] We focus on transformations that are observable by the econometrician, in which case Γ_{jt} will serve as the instrumental variables for the problem.[16]

One challenge in using equation (7) for identification is the presence of an endogenous variable m_{jt} in the model that is correlated with η_{jt}. However, all lagged output/input values, as well as the current values of the predetermined inputs k_{jt} and l_{jt}, are transformations of $\mathcal{I}_{jt-1}$.[17] Therefore, the full vector of potential instrumental variables given the data described in section II.A is given by $\Gamma_{jt} = (k_{jt}, l_{jt}, d_{t-1}, y_{jt-1}, k_{jt-1}, l_{jt-1}, m_{jt-1}, \ldots, d_1, y_{j1}, k_{j1}, l_{j1}, m_{j1})$.[18]

A. *Identification*

Despite the apparent abundance of available instruments for the flexible input m_{jt}, notice that by replacing for ω_{jt} in the intermediate-input demand equation (5), we obtain

$$m_{jt} = \mathbb{M}\left(k_{jt}, l_{jt}, h\left(\mathbb{M}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1}\right) + \eta_{jt} - d_t\right). \quad (8)$$

This implies that the only sources of variation left in m_{jt} after conditioning on $\left(k_{jt}, l_{jt}, d_{t-1}, k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \in \Gamma_{jt}$ (which are used as instruments for themselves) are the unobservable η_{jt} itself and d_t. Therefore, for all of the remaining elements in Γ_{jt}, their only power as instruments is via their dependence on d_t.

[15] Notice that since ε_{jt} is recoverable from the first stage, one could instead use the orthogonality $E[\eta_{jt} \mid \Gamma_{jt}] = 0$. However, this can be formed only for observations in which the proxy variable—intermediate-input demand (or investment in Olley and Pakes 1996)—is strictly positive. Observations that violate the strict monotonicity of the proxy equation need to be dropped from the first stage, which implies that ε_{jt} cannot be recovered. This introduces a selection bias since $E[\eta_{jt} \mid \Gamma_{jt}, \iota_{jt} > 0] \neq E[\eta_{jt} \mid \Gamma_{jt}]$, where ι_{jt} is the proxy variable. The reason is that firms that receive lower draws of η_{jt} are more likely to choose nonpositive values of the proxy, and this probability is a function of the other state variables of the firm.

[16] The idea that one can use expectations conditional on lagged information sets to exploit the property that the innovation should be uncorrelated with lagged variables goes back to at least the work on rational expectations models; see, e.g., Sargent (1978) and Hansen and Sargent (1980).

[17] If k_{jt} and/or l_{jt} are dynamic but not predetermined, then only lagged values enter Γ_{jt}.

[18] Following Doraszelski and Jaumandreu (2013), we exclude d_t from the instruments, as current prices and the innovation to productivity are determined contemporaneously and hence may be correlated (see also Ackerberg et al. 2007).

Identification of the production function f by instrumental variables is based on projecting output y_{jt} onto the exogenous variables Γ_{jt} (see, e.g., Newey and Powell 2003). This generates a restriction between (f, h) and the distribution of the data that takes the form

$$\begin{aligned} E[y_{jt} \mid \Gamma_{jt}] &= E[f(k_{jt}, l_{jt}, m_{jt}) \mid \Gamma_{jt}] + E[\omega_{jt} \mid \Gamma_{jt}] \\ &= E[f(k_{jt}, l_{jt}, m_{jt}) \mid \Gamma_{jt}] + h(\phi(k_{jt-1}, l_{jt-1}, m_{jt-1}) \\ &\quad + d_{t-1} - f(k_{jt-1}, l_{jt-1}, m_{jt-1})). \end{aligned} \tag{9}$$

The unknown functions underlying equation (9) are given by (f, h), since $\phi(k_{jt-1}, l_{jt-1}, m_{jt-1}) + d_{t-1}$ is known from the first-stage equation (6). The true (f^0, h^0) are identified if no other $(\tilde{f}, \tilde{h})$ among all possible alternatives also satisfy the functional restriction (9) given the distribution of the observables.[19]

In theorem 1, we first show that in the absence of time-series variation in prices, $d_t = d \;\forall\; t$, the proxy variable structure does not suffice to identify the gross output production function.[20] Specifically, we show that the application of instrumental variables (via the orthogonality restriction $E[\eta_{jt} + \varepsilon_{jt} \mid \Gamma_{jt}] = 0$) to equation (7) is insufficient for identifying the production function f (and the Markovian process h). Intuitively, if d_t does not vary over time in equation (8), then the only remaining source of variation in m_{jt} is the innovation η_{jt}, which is by construction orthogonal to the remaining elements of Γ_{jt}.

Theorem 1. *In the absence of time-series variation in relative prices, $d_t = d \;\forall\; t$, under the model defined by assumptions 1–4, there exists a continuum of alternative $(\tilde{f}, \tilde{h})$ defined by*

$$\begin{aligned} \tilde{f}(k_{jt}, l_{jt}, m_{jt}) &\equiv (1 - a)f^0(k_{jt}, l_{jt}, m_{jt}) + a\phi(k_{jt}, l_{jt}, m_{jt}), \\ \tilde{h}(x) &\equiv ad + (1 - a)h^0\left(\frac{1}{(1 - a)}(x - ad)\right) \end{aligned}$$

for any $a \in (0, 1)$, that satisfies the same functional restriction (9) as the true (f^0, h^0).

Proof. We begin by noting that from the definition of ϕ, it follows that $E[y_{jt} \mid \Gamma_{jt}] = E[\phi(k_{jt}, l_{jt}, m_{jt}) + d_t \mid \Gamma_{jt}]$. Hence, for any (f, h) that satisfy (9), it must be the case that

[19] Some researchers may not be interested in recovering h. In our results below, regardless of whether h is identified, the production function f is not (except in the degenerate case in which there are no differences in ω across firms, so $\phi(k_{jt}, l_{jt}, m_{jt}) = f(k_{jt}, l_{jt}, m_{jt})$).

[20] In app. O1, we show that a similar result holds for the case of investment as the proxy variable and for the use of dynamic panel techniques under this same structure.

$$\begin{aligned} &E\left[\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t - f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] \\ &\quad = h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + \ d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{aligned} \tag{10}$$

Next, given the definition of $(\tilde{f}, \tilde{h})$ and noting that $d_t = d \ \forall \ t$, we have

$$\begin{aligned} &\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ \tilde{h}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\left(\phi\left(k_{jt}, l_{jt}, m_{jt}\right) - f^0\left(k_{jt}, l_{jt}, m_{jt}\right)\right) + ad \\ &\qquad + (1-a)h^0\left(\frac{(1-a)\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right)}{1-a}\right) \\ &\quad = f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a(\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d - f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \\ &\qquad + (1-a)h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{aligned}$$

Now, take the conditional expectation of the above (with respect to Γ_{jt}):

$$\begin{aligned} &E\left[\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + \ \tilde{h}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = E\left[f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + ah^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right. \\ &\qquad \left. + \ d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\qquad + (1-a)h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &\quad = E\left[f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}\right] + h^0\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right. \\ &\qquad \left. + \ d - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right), \end{aligned}$$

where the first equality uses the relation in equation (10). Thus, (f^0, h^0) and $(\tilde{f}, \tilde{h})$ satisfy the functional restriction (9) and cannot be distinguished via instrumental variables. QED

We now provide two corollaries to our main theorem to describe the extent to which time-series variation (via d_t) can be used to identify the model. (In app. O2, we provide an illustration of these results in the context of the commonly employed Cobb-Douglas parametric form.)

In corollary 1, we show that if $T = 2$ (the minimum number of periods required by these procedures), the model cannot be identified, even if d_t varies. Intuitively, since the second stage already conditions on d_1, the only remaining potential source of variation is in d_2, which of course does not vary.

Corollary 1. For $T = 2$, under the model defined by assumptions 1–4, there exists a continuum of alternative $(\tilde{f}, \tilde{h})$ defined by

$$\begin{aligned} \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) &\equiv (1-a)f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\phi\left(k_{jt}, l_{jt}, m_{jt}\right), \\ \tilde{h}(x) &\equiv ad_2 + (1-a)h^0\left(\frac{1}{(1-a)}(x - ad_1)\right) \end{aligned}$$

for $t = 1, 2$ and for any $a \in (0, 1)$ that satisfies the same functional restriction (9) as the true (f^0, h^0).

Proof. The proof follows from the same steps in the proof of theorem 1. QED

In corollary 2, we show that when one relaxes the assumption of time homogeneity in either the production function or the Markov process for productivity, the model similarly cannot be identified, even with $T > 2$. Intuitively, once the model varies with time, time-series variation is no longer helpful.

Corollary 2. Under the model defined by assumptions 1–4,

i) if the production function is time varying, f_t^0, there exists a continuum of alternative $(\tilde{f}_t, \tilde{h})$ defined by[21]

$$\tilde{f}_t(k_{jt}, l_{jt}, m_{jt}) \equiv (1 - a)f_t^0(k_{jt}, l_{jt}, m_{jt}) + a\phi_t(k_{jt}, l_{jt}, m_{jt}) + ad_t,$$
$$\tilde{h}(x) \equiv (1 - a)h^0\left(\frac{1}{(1 - a)}x\right),$$

or

ii) if the process for productivity is time varying, h_t^0, there exists a continuum of alternative $(\tilde{f}, \tilde{h}_t)$ defined by

$$\tilde{f}(k_{jt}, l_{jt}, m_{jt}) \equiv (1 - a)f^0(k_{jt}, l_{jt}, m_{jt}) + a\phi(k_{jt}, l_{jt}, m_{jt}),$$
$$\tilde{h}_t(x) \equiv ad_t + (1 - a)h_t^0\left(\frac{1}{(1 - a)}(x - ad_{t-1})\right),$$

such that for any $a \in (0, 1)$, these alternative functions satisfy the functional restriction (9).

Proof. The proof follows from the same steps in the proof of theorem 1. QED

The result in theorem 1 and its two corollaries is a useful benchmark, as it directly relates to the econometric approach used in the proxy variable literature. However, this instrumental variables approach does not necessarily exhaust the sources of identification inherent in the proxy variable structure. First, since the instrumental variables approach is based only on conditional expectations, it does not employ the entire distribution of the data $(y_{jt}, m_{jt}, \Gamma_{jt})$. Second, it does not directly account for the fact that assumption 3 also imposes restrictions (scalar unobservability and monotonicity) on the determination of the endogenous variable m_{jt} via $\mathbb{M}(\cdot)$. Therefore, the proxy variable structure imposes restrictions on a simultaneous system of equations because, in addition to the

[21] Notice that when the production function is allowed to be time varying, the first-stage estimates also need to be time varying (i.e., $E[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}] = \phi_t(k_{jt}, l_{jt}, m_{jt}) + d_t$).

model for output, y_{jt}, there is a model for the proxy variable—in this case, intermediate inputs, m_{jt}. In appendix O3, we extend our result to the full model involving f, h, and $\mathbb{M}$, using the full distribution of the data.

B. Monte Carlo Evidence on the Use of Time-Series Variation

The result in theorem 1 shows that under the model described above, there are not enough sources of cross-sectional variation that can be used to identify the gross output production function. In particular, the problem is associated with flexible intermediate inputs. While aggregate time-series variation provides a potential source of identification, relying on it runs a risk of weak identification in practice.

To evaluate the performance of using time-series variation as a source of identification, we conduct several Monte Carlo experiments. As we show in equation (5), the firm's optimal choice of intermediate inputs depends on the relative price of intermediate inputs to output, as opposed to the levels. In our simulations, we fix the price of output to be one and let the price of intermediate inputs vary. Specifically, the (log) price of intermediate inputs is assumed to follow an AR(1) (first-order autoregressive) process. We refer to the variance of the innovation in this process as the level of time-series variation.

The parameters of the data-generating process are chosen to roughly match the properties of our data, as well as the variances of our productivity estimates, described in section VII. A full description of the setup is provided in appendix O4 (Monte Carlo 1). The key features are as follows. For simplicity, we abstract away from labor and specify a Cobb-Douglas production function in capital and intermediate inputs, with elasticities of 0.25 and 0.65, respectively. Firms maximize the expected stream of future discounted profits. Productivity is assumed to evolve according to an AR(1) process with a persistence parameter of 0.8. The law of motion for capital is given by $K_{jt} = (1 - \kappa_j)K_{jt-1} + I_{jt-1}$, where investment I is chosen a period ahead in $t - 1$ and the depreciation rate $\kappa_j \in [0.05, 0.15]$ varies across firms. Intermediate inputs are chosen flexibly in period t as a function of capital, productivity, and the relative price of intermediates to output. The price of investment is assumed to be fixed. For the time-series process for the price of intermediate inputs, we set the AR(1) coefficient to 0.6 and the variance of the innovation at a baseline value of 0.0001.[22] In addition to the baseline value of time-series variation, we also create versions with half, twice, and 10 times this baseline variation (0.00005, 0.0002, and 0.001, respectively).

[22] This corresponds to the values obtained from a regression of the log relative price of intermediate inputs on its lag for the largest industry in Chile: food products (International Standard Industrial Classification [ISIC] code 311). The level of time-series variation in Colombia is considerably smaller.

We construct 12 different panel structures: 200 versus 500 firms and 3, 5, 10, 20, 30, and 50 periods. For each panel and for each of the four different levels of time-series variation, we simulate 500 data sets. We estimate a version of the proxy variable technique applied to gross output, as described above, using intermediate inputs as the proxy. To reduce the potential noise from nonparametric estimation, we impose the true parametric structure of the model in the estimation routine (i.e., a Cobb-Douglas production function and an AR(1) process for productivity).

As the results in figure 1 illustrate, even with twice the baseline level of time-series variation and even with very long panels (50 periods), the proxy variable technique applied to gross output consistently generates significantly biased estimates of the production function. It is only when we boost the level of aggregate variation to 10 times the baseline obtained from the data and for relatively large panels that the estimates start to converge to the truth. However, even in this case, the 2.5%–97.5% interquantile range of the estimates is quite wide (see fig. O4.1; figs. O4.1–O4.5 are available online).

The results described above show that using time-series variation as a source of identification, while valid in theory, may not perform well in practice. However, it also suggests that if there were observed shifters that varied across firms, which entered the flexible input demand $\mathbb{M}$ but were excluded from the production function, then this additional variation could be used to better identify the production function.[23] In particular, firm-varying flexible input and output prices are one source of such variation that has been considered recently by Doraszelski and Jaumandreu (2013, 2018), which we discuss in more detail in section VI. In the next section, we develop an alternative identification and estimation strategy that does not rely on researchers having access to long panels with rich aggregate time-series variation or additional sources of exogenous cross-sectional variation, such as firm-specific prices.

IV. Nonparametric Identification via First-Order Conditions

In this section, we show that the restrictions implied by the optimizing behavior of the firm, combined with the idea of using lagged inputs as instruments employed by the dynamic panel and proxy variable literatures, are sufficient to nonparametrically identify the production function

[23] It may be possible to achieve identification in the absence of exclusion restrictions by imposing additional restrictions (see Koopmans, Rubin, and Leipnik 1950; Heckman and Robb 1985). One example is using heteroskedasticity restrictions (see, e.g., Rigobon 2003; Klein and Vella 2010; Lewbel 2012), although these approaches require explicit restrictions on the form of the error structure. We thank an anonymous referee for pointing this out. We are not aware of any applications of these ideas in the production function setting.

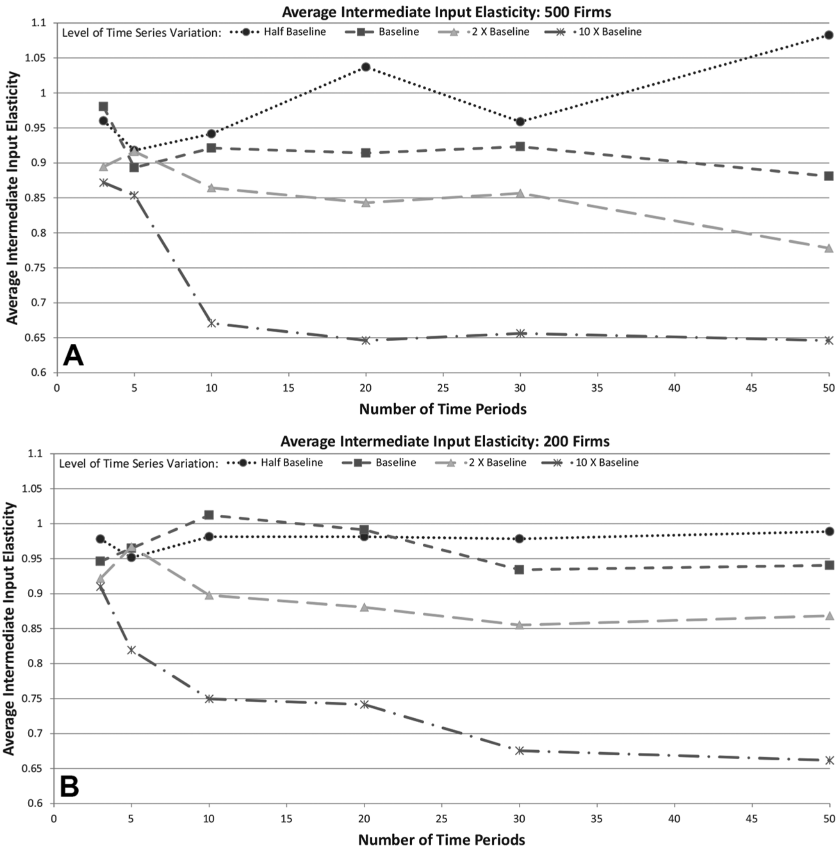

FIG. 1.—Monte Carlo–proxy variable estimator applied to gross output. This figure presents the results from applying a proxy variable estimator extended to gross output to Monte Carlo data generated as described in section III and appendix O4. The data are generated under four different levels of time-series variation. The *X*-axis measures the number of time periods in the panel used to generate the data. The *Y*-axis measures the average of the estimated intermediate-input elasticity across 100 Monte Carlo simulations. The true value of the elasticity is 0.65. Panel *A* includes 500 firms; panel *B* includes 200 firms. A color version of this figure is available online.

and productivity, even absent additional sources of exogenous variation in flexible inputs.[24] The key idea is to recognize that the production function and the intermediate-input demand, f and $\mathbb{M}$, are not independent functions for an optimizing firm. The input demand $\mathbb{M}$ is implicitly defined by f through the firm's first-order condition. This connection generates cross-equation restrictions that have been recognized and exploited in parametric settings (for early examples, see Klein 1953; Solow 1957; Nerlove 1963;

[24] Please see app. O6 for the extension to the case of multiple flexible inputs.

for more recent examples, see Doraszelski and Jaumandreu 2013, 2018).[25] Our contribution is to show that this functional relationship can be exploited in a fully nonparametric fashion to nonparametrically identify the entire production function. The reason why we are able to use the first-order condition with such generality is that the proxy variable assumption—assumption 3—already presumes that intermediate inputs are a flexible input, thus making the economics of this input choice especially tractable.

The first step of our identification strategy is to recognize the nonparametric link between the production function (1) and the first-order condition (4). Taking logs of (4) and differencing with the production function gives

$$\begin{aligned} s_{jt} &= \ln \mathcal{E} + \ln\left(\frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right)\right) - \varepsilon_{jt} \\ &\equiv \ln\ D^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}\right) - \varepsilon_{jt}, \end{aligned} \tag{11}$$

where $s_{jt} \equiv \ln((\rho_t M_{jt})/(P_t Y_{jt}))$ represents the (log) intermediate-input share of output. In the following theorem, we prove that, since $E[\varepsilon_{jt} \mid k_{jt}, l_{jt}, m_{jt}] = 0$, both the output elasticity of the flexible input and ε_{jt} can be recovered by regressing the shares of intermediate inputs s_{jt} on the vector of inputs (k_{jt}, l_{jt}, m_{jt}).

Theorem 2. Under assumptions 1–4, and assuming that ρ_t/P_t (or the relative price deflator) is observed, the share regression in equation (11) nonparametrically identifies the flexible input elasticity $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$ of the production function almost everywhere in (k_{jt}, l_{jt}, m_{jt}).

Proof. Given the flexible input demand $m_{jt} = \mathbb{M}_t(k_{jt}, l_{jt}, \omega_{jt})$, and since $k_{jt}, l_{jt}, \omega_{jt} \in \mathcal{I}_{jt}$, assumption 2 implies that $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}, k_{jt}, l_{jt}, m_t] = E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$. Hence, $E[\varepsilon_{jt} \mid k_{jt}, l_{jt}, m_{jt}] = 0$ by the law of iterated expectations. As a consequence, the conditional expectation based on equation (11),

$$E\left[s_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = \ln\ D^{\mathcal{E}}(k_{jt}, l_{jt}, m_{jt}), \tag{12}$$

identifies the function $D^{\mathcal{E}}$. We refer to this regression in the data as the *share* regression.

Observe that $\varepsilon_{jt} = \ln\ D^{\mathcal{E}}(k_{jt}, l_{jt}, m_{jt}) - s_{jt}$ and thus the constant

$$\mathcal{E} = E\left[\exp\left(\ln\ D^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}\right) - s_{jt}\right)\right] \tag{13}$$

can be identified.[26] This allows us to identify the flexible input elasticity as

[25] See sec. VI for a more detailed discussion of this literature.

[26] Doraszelski and Jaumandreu (2013) suggest using this approach to identify this constant in the context of a Cobb-Douglas production function.

$$D\left(k_{jt}, l_{jt}, m_{jt}\right) \equiv \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) = \frac{D^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}\right)}{\mathcal{E}}. \tag{14}$$

QED

Theorem 2 shows that, by taking full advantage of the economic content of the model, we can identify the flexible input elasticity using moments in ε_{jt} alone. The theorem is written under the assumption that $P_{\varepsilon}(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_{\varepsilon}(\varepsilon_{jt})$ (in assumption 2). Much of the previous literature assumes only mean independence $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$. As with the proxy variable methods, our approach can be adapted to work under the weaker mean independence assumption as well. In this case, we would have that, from the firm's problem, $\mathcal{E}(\mathcal{I}_{jt}) \equiv E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}]$. Since ε_{jt} (and hence $e^{\varepsilon_{jt}}$) is identified from the share regression (12), $\mathcal{E}(\mathcal{I}_{jt})$ can also be identified. In terms of the proof, the elasticity would then be obtained as $D(k_{jt}, l_{jt}, m_{jt}) = [D^{\mathcal{E}}(k_{jt}, l_{jt}, m_{jt}, \mathcal{I}_{jt})/\mathcal{E}(\mathcal{I}_{jt})]$, where we notice that now $D^{\mathcal{E}}(k_{jt}, l_{jt}, m_{jt}, \mathcal{I}_{jt})$ depends on $\mathcal{I}_{jt}$ and hence the share regression would need to be adjusted accordingly.[27]

The next step in our approach is to use the information from the share regression to recover the rest of the production function nonparametrically. The idea is that the flexible input elasticity defines a partial differential equation that can be integrated up to identify the part of the production function f related to the intermediate input m.[28] By the fundamental theorem of calculus, we have

$$\int \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) dm_{jt} = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathcal{C}\left(k_{jt}, l_{jt}\right). \tag{15}$$

Subtracting equation (15) from the production function and rearranging terms, we have

$$\mathcal{Y}_{jt} \equiv y_{jt} - \varepsilon_{jt} - \int \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) dm_{jt} = -\mathcal{C}\left(k_{jt}, l_{jt}\right) + \omega_{jt}. \tag{16}$$

Notice that $\mathcal{Y}_{jt}$ is an "observable" random variable, as it is a function of data, as well as the flexible input elasticity and the ex post shock, which are recovered from the share regression.

We then follow the dynamic panel literature (as well as the proxy variable literature) and use the Markovian structure on productivity in assumption 2 to generate moments based on the panel structure of the data and recover $\mathcal{C}(k_{jt}, l_{jt})$. By replacing for ω_{jt} in equation (16), we have

[27] In practice, conditioning on the entire information set is infeasible, but one could include only a rolling subset of $\mathcal{I}_{jt}$—e.g., (k_{jt}, l_{jt})—and recover $\mathcal{E}(\mathcal{I}_{jt})$ by running a nonparametric regression of $e^{\varepsilon_{jt}}$ on the relevant elements of $\mathcal{I}_{jt}$.

[28] See Houthakker (1950) for the related problem of how to recover the utility function from the demand functions.

$$\mathcal{Y}_{jt} = -\mathcal{C}(k_{jt}, l_{jt}) + h(\mathcal{Y}_{jt-1} + \mathcal{C}(k_{jt-1}, l_{jt-1})) + \eta_{jt}. \tag{17}$$

Since $(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ are all known to the firm at period $t - 1$ and (k_{jt}, l_{jt}) are predetermined, we have the orthogonality $E[\eta_{jt} \mid k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}] = 0$, which implies that

$$E[\mathcal{Y}_{jt} \mid k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}] = -\mathcal{C}(k_{jt}, l_{jt}) + h(\mathcal{Y}_{jt-1} + \mathcal{C}(k_{jt-1}, l_{jt-1})). \tag{18}$$

A regression of $\mathcal{Y}_{jt}$ on $(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ identifies the left-hand side of equation (18). Intuitively, if one can vary (k_{jt}, l_{jt}) separately from $(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ for all points in the support of (k_{jt}, l_{jt}), then $\mathcal{C}$ can be separately identified from h up to an additive constant.[29]

We now establish this result formally in theorem 3 on the basis of the above discussion. To do so, we first formalize the support condition described in the paragraph above in the following regularity condition on the support of the regressors $(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}$; adapted from Newey, Powell, and Vella 1999).

Assumption 5. For each point $(\bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1})$ in the support of $(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$, the boundary of the support of (k_{jt}, l_{jt}) conditional on $(\bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1})$ has a probability measure zero.

Assumption 5 is a condition that states that we can independently vary the predetermined inputs (k_{jt}, l_{jt}) conditional on $(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ within the support. This implicitly assumes the existence of enough variation in the input demand functions for the predetermined inputs to induce open set variation in them conditional on the lagged output and input values $(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$. This condition makes explicit the variation that allows for nonparametric identification of the remainder of the production function under the second-stage moments above. A version of this assumption is thus implicit in the second stage of the proxy variable procedures. Note that this assumption rules out mass points in the boundary of the support, which may arise from discrete decisions such as entry and exit. In footnote 30, we discuss how one can still identify the production function if this is the case, under a mild additional restriction.

Theorem 3. Under assumptions 1–5, if $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$ is nonparametrically known, then the production function f is nonparametrically identified up to an additive constant.

Proof. Assumptions 2, 3, and 5 ensure that with probability one for any (k_{jt}, l_{jt}, m_{jt}) in the support of the data there is a set

$$\{(k, l, m) \mid k = k_{jt}, l = l_{jt}, m \in [m\,(k_{jt}, l_{jt}), m_{jt}]\}$$

also contained in the support for some $m(k_{jt}, l_{jt})$. Hence, with probability one the integral

[29] As it is well known, a constant in the production function and mean productivity, $E[\omega_{jt}]$, are not separately identified.

$$\int_{m\left(k_{jt},l_{jt}\right)}^{m_{jt}} \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m\right) dm = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathcal{C}(k_{jt}, l_{jt})$$

is identified, where the equality follows from the fundamental theorem of calculus. Therefore, if two production functions f and $\tilde{f}$ give rise to the same input elasticity $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$ over the support of the data, then they can differ only by an additive function $\mathcal{C}(k_{jt}, l_{jt})$. To identify this additive function, observe that we can identify the joint distribution of $(\mathcal{Y}_{jt}, k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ for $\mathcal{Y}_{jt}$ defined by (16). Thus, the regression function

$$E\left[\mathcal{Y}_{jt} \mid k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}\right] = \mu(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}) \qquad (19)$$

can be identified for almost all $x_{jt} = (k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$, where, given equation (18), $\mu(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}) = -\mathcal{C}(k_{jt}, l_{jt}) + h(\mathcal{Y}_{jt-1} + \mathcal{C}(k_{jt-1}, l_{jt-1}))$. Let $(\tilde{\mathcal{C}}, \tilde{h})$ be an alternative pair of functions; $(\mathcal{C}, h)$ and $(\tilde{\mathcal{C}}, \tilde{h})$ are observationally equivalent if and only if

$$\begin{aligned} &-\mathcal{C}\left(k_{jt}, l_{jt}\right) + h\left(\mathcal{Y}_{jt-1} + \mathcal{C}\left(k_{jt-1}, l_{jt-1}\right)\right) \\ &= -\tilde{\mathcal{C}}\left(k_{jt}, l_{jt}\right) + \tilde{h}\left(\mathcal{Y}_{jt-1} + \tilde{\mathcal{C}}\left(k_{jt-1}, l_{jt-1}\right)\right) \end{aligned} \qquad (20)$$

for almost all points in the support of x_{jt}. Our support assumption (assumption 5) on (k_{jt}, l_{jt}) allows us to take partial derivatives of both sides of (20) with respect to k_{jt} and l_{jt}:

$$\frac{\partial}{\partial z}\mathcal{C}\left(k_{jt}, l_{jt}\right) = \frac{\partial}{\partial z}\tilde{\mathcal{C}}(k_{jt}, l_{jt})$$

for $z \in \{k_{jt}, l_{jt}\}$ and for all x_{jt} in its support, which implies that $\mathcal{C}(k_{jt}, l_{jt}) - \tilde{\mathcal{C}}(k_{jt}, l_{jt}) = c$ for a constant c for almost all x_{jt}.[30] Thus, we have shown that the production function is identified up to a constant. QED

Theorem 3 demonstrates that if one can recover the elasticity of the flexible input, as we do via the share regression, the production function is nonparametrically identified. This result highlights the importance of recognizing the nonparametric link between the production function and the first-order condition of the firm that allowed us to recover the flexible elasticity in the first place. It also demonstrates the power of dynamic panel methods under a (typically implicit) rank condition such

[30] Assumption 5 rules out mass points in the boundary of the support of (k_{jt}, l_{jt}) conditional on $(\bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1})$. However, even if such mass points exist, the steps of the proof above show that one can identify $\mathcal{C}(k_{jt}, l_{jt})$ everywhere else (besides the mass points) up to a constant c. To identify $\mathcal{C}(k_{jt}, l_{jt})$ at the mass points, consider a mass point (k_{jt}^{*}, l_{jt}^{*}) conditional on $(\bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1})$. As long as there exists a point (k_{jt}', l_{jt}') in the interior of the support conditional on $(\bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1})$, we can construct the unknown $\mathcal{C}(k_{jt}^{*}, l_{jt}^{*})$ as

$$\mathcal{C}\left(k_{jt}^{*}, l_{jt}^{*}\right) = -E\left[\mathcal{Y}_{jt} \mid k_{jt}^{*}, l_{jt}^{*}, \bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1}\right] + E\left[\mathcal{Y}_{jt} \mid k_{jt}', l_{jt}', \bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1}\right] + \mathcal{C}\left(k_{jt}', l_{jt}'\right),$$

again up to the constant c.

as assumption 5. Under this rank condition, if there were no flexible inputs and ε were known, one could nonparametrically identify the gross output production function (and productivity) on the basis of dynamic panel methods alone. We revisit this in our discussion of dynamic panel methods in section VI.C.

Our results in theorems 2 and 3 are derived under the assumption that the model structure is time invariant. It is straightforward to generalize them to the time-varying case by indexing the production function f and the Markov process h by time t, simply repeating the steps of our analysis separately for each time period $t \in \{2, \ldots, T\}$.

V. A Computationally Simple Estimator

In this section, we show how to obtain a simple nonparametric estimator of the production function using standard sieve series estimators as analyzed by Chen (2007). Our estimation procedure consists of two steps. We first show how to estimate the share regression and then proceed to estimation of the constant of integration $\mathcal{C}$ and the Markov process h.

We propose a finite-dimensional truncated linear series given by a complete polynomial of degree r for the share regression. Given the observations $\{(y_{jt}, k_{jt}, l_{jt}, m_{jt})\}_{t=1}^{T}$ for the firms $j = 1, \ldots, J$ sampled in the data, we propose to use a complete polynomial of degree r in k_{jt}, l_{jt}, m_{jt} and to use the sum of squared residuals, $\Sigma_{jt}\varepsilon^2_{jt}$, as our objective function. For example, for a complete polynomial of degree two, our estimator would solve

$$\min_{\gamma'} \sum_{j,t} \left\{ s_{jt} - \ln \begin{pmatrix} \gamma'_0 + \gamma'_k k_{jt} + \gamma'_l l_{jt} + \gamma'_m m_{jt} + \gamma'_{kk} k^2_{jt} + \gamma'_{ll} l^2_{jt} \\ + \gamma'_{mm} m^2_{jt} + \gamma'_{kl} k_{jt} l_{jt} + \gamma'_{km} k_{jt} m_{jt} + \gamma'_{lm} l_{jt} m_{jt} \end{pmatrix} \right\}^2 .$$

The solution to this problem is an estimator

$$D^{\mathcal{E}}_r\left(k_{jt}, l_{jt}, m_{jt}\right) = \sum_{r_k + r_l + r_m \leq r} \gamma'_{r_k, r_l, r_m} k^{r_k}_{jt} l^{r_l}_{jt} m^{r_m}_{jt} \text{ with } r_k, r_l, r_m \geq 0 \qquad (21)$$

of the elasticity up to the constant $\mathcal{E}$, as well as the residual ε_{jt} corresponding to the ex post shocks to production.[31] Since we can estimate $\hat{\mathcal{E}} = (1/(JT))\Sigma_{j,t} e^{\hat{\varepsilon}_{jt}}$, we can recover $\hat{\gamma} \equiv \hat{\gamma}'/\hat{\mathcal{E}}$ and thus estimate $(\partial f(k_{jt}, l_{jt}, m_{jt})/\partial m_{jt})$ from equation (21), free of this constant.

Given our estimator for the intermediate-input elasticity, we can calculate the integral in (15). One advantage of the polynomial sieve estimator we use is that this integral will have a closed-form solution:

[31] As with all nonparametric sieve estimators, the number of terms in the series increases with the number of observations. Under mild regularity conditions, these estimators will be consistent and asymptotically normal for sieve M-estimators such as the one we propose. See Chen (2007).

$$\mathcal{D}_r\left(k_{jt}, l_{jt}, m_{jt}\right) \equiv \int D_r\left(k_{jt}, l_{jt}, m_{jt}\right) dm_{jt} = \sum_{r_k + r_l + r_m \leq r} \frac{\gamma_{r_k, r_l, r_m}}{r_m + 1} k_{jt}^{r_k} l_{jt}^{r_l} m_{jt}^{r_m + 1}.$$

For a degree two estimator ($r = 2$), we would have

$$\mathcal{D}_2\left(k_{jt}, l_{jt}, m_{jt}\right) \equiv \left(\begin{array}{c} \gamma_0 + \gamma_k k_{jt} + \gamma_l l_{jt} + \dfrac{\gamma_m}{2} m_{jt} + \gamma_{kk} k_{jt}^2 + \gamma_{ll} l_{jt}^2 \\ + \dfrac{\gamma_{mm}}{3} m_{jt}^2 + \gamma_{kl} k_{jt} l_{jt} + \dfrac{\gamma_{km}}{2} k_{jt} m_{jt} + \dfrac{\gamma_{lm}}{2} l_{jt} m_{jt} \end{array} \right) m_{jt}.$$

With an estimate of ε_{jt} and $\mathcal{D}_r(k_{jt}, l_{jt}, m_{jt})$ in hand, we can form a sample analogue of $\mathcal{Y}_{jt}$ in equation (16): $\hat{\mathcal{Y}}_{jt} \equiv \ln(Y_{jt}/(e^{\hat{\varepsilon}_{jt}} e^{\hat{\mathcal{D}}_r(k_{jt}, l_{jt}, m_{jt})}))$.

In the second step, to recover the constant of integration $\mathcal{C}$ in (17) and the Markovian process h, we use similar complete polynomial series estimators. Since a constant in the production function cannot be separately identified from mean productivity, $E[\omega_{jt}]$, we normalize $\mathcal{C}(k_{jt}, l_{jt})$ to contain no constant. That is, we use

$$\mathcal{C}_\tau\left(k_{jt}, l_{jt}\right) = \sum_{0 < \tau_k + \tau_l \leq \tau} \alpha_{\tau_k, \tau_l} k_{jt}^{\tau_k} l_{jt}^{\tau_l} \text{ with } \tau_k, \tau_l \geq 0 \tag{22}$$

and

$$h_A\left(\omega_{jt-1}\right) = \sum_{0 \leq a \leq A} \delta_a \omega_{jt-1}^a \tag{23}$$

for some degrees τ and A (that increase with the sample size). Combining these and replacing for ω_{jt-1}, we have the estimating equation

$$\hat{\mathcal{Y}}_{jt} = - \sum_{0 < \tau_k + \tau_l \leq \tau} \alpha_{\tau_k, \tau_l} k_{jt}^{\tau_k} l_{jt}^{\tau_l} + \sum_{0 \leq a \leq A} \delta_a \left(\hat{\mathcal{Y}}_{jt-1} + \sum_{0 < \tau_k + \tau_l \leq \tau} \alpha_{\tau_k, \tau_l} k_{jt-1}^{\tau_k} l_{jt-1}^{\tau_l} \right)^a + \eta_{jt}. \tag{24}$$

We can then use moments of the form $E[\eta_{jt} k_{jt}^{\tau_k} l_{jt}^{\tau_l}] = 0$ and $E[\eta_{jt} \hat{\mathcal{Y}}_{jt-1}^a] = 0$ to form a standard sieve moment criterion function to estimate (α, δ).[32] Putting the two stages together, we have the following moments:

$$\begin{aligned} E\left[\varepsilon_{jt} \frac{\partial \ln\ D_r\left(k_{jt}, l_{jt}, m_{jt}\right)}{\partial \gamma}\right] &= 0, \\ E\left[\eta_{jt} k_{jt}^{\tau_k} l_{jt}^{\tau_l}\right] &= 0, \\ E\left[\eta_{jt} \mathcal{Y}_{jt-1}^a\right] &= 0, \end{aligned}$$

where the first set of moments shows the nonlinear least squares moments corresponding to the share equation.

[32] Alternatively, for a guess of α, one can form $\omega_{jt-1}(\alpha) = \hat{\mathcal{Y}}_{jt-1} + \mathcal{C}(k_{jt-1}, l_{jt-1}) = \hat{\mathcal{Y}}_{jt-1} + \Sigma_{0 < \tau_k + \tau_l \leq \tau} \alpha_{\tau_k, \tau_l} k_{jt-1}^{\tau_k} l_{jt-1}^{\tau_l}$ and use moments of the form $E[\eta_{jt} \omega_{jt-1}(\alpha)] = 0$ to estimate δ. Notice that since $\omega_{jt}(\alpha) = \hat{\mathcal{Y}}_{jt} + \Sigma_{0 < \tau_k + \tau_l \leq \tau} \alpha_{\tau_k, \tau_l} k_{jt}^{\tau_k} l_{jt}^{\tau_l}$, this is equivalent to regressing ω_{jt} on a sieve in ω_{jt-1}. Then the moments $E[\eta_{jt} k_{jt}^{\tau_k} l_{jt}^{\tau_l}] = 0$ can be used to estimate α.

Under the just-identified case described above, our two-step sieve procedure is a sieve M-estimator. As we show in appendix O5, our setting can be mapped to the setting in Hahn, Liao, and Ridder (2018), where they establish the consistency and asymptotic normality of plug-in, two-step sieve estimators such as the one we employ. Furthermore, they show that the asymptotic variance of this estimator is numerically equivalent to the parametric two-step version. Therefore, we can apply their numerical equivalence results and conduct inference as if our sieves were the true parametric structure.[33] To compute standard errors for the functionals of interest (e.g., elasticities), we employ a nonparametric bootstrap (see, e.g., Horowitz 2001).[34]

VI. Relationship to Literature

A. *Price Variation as an Instrument*

Recall that theorem 1 (and its extensions in apps. O1 and O3) shows that, absent additional sources of variation, dynamic panel/proxy variable methods cannot be used to identify the gross output production function. As discussed in section III, cross-sectional variation in prices can potentially be used to identify the production function by providing a source of variation for flexible inputs. The literature, however, has identified several challenges to using prices as instruments (see Griliches and Mairesse 1998; Ackerberg et al. 2007). First, in many firm-level production data sets, firm-specific prices are simply not observed. Second, even if price variation is observed, to be useful as an instrument, the variation employed must not be correlated with the innovation to productivity, η_{jt}, and it cannot solely reflect differences in the quality of either inputs or output. To the extent that input and output prices capture quality differences, prices should be included in the measure of the inputs used in production.[35]

This is not to say that if one can isolate exogenous price variation (e.g., if prices vary because of segmented geographic markets or because of policy shocks), it cannot be used to aid in identification. The point is that simply observing price variation is not enough. The case must be made that the price variation that is used is indeed exogenous. For example, if prices are observed and serially correlated, one way to deal with the endogeneity concern, as suggested by Doraszelski and Jaumandreu (2013), is to use

[33] One could also use lags of inputs to estimate an overidentified version of the model. In this case, the second stage of our estimator becomes a sieve MD-estimator. We are not aware of any similar numerical equivalence results for such estimators.

[34] In app. O4 (Monte Carlo 4), we present Monte Carlo simulations that show that our bootstrap procedure has the correct coverage for the nonparametric estimates.

[35] Recent work has suggested that quality differences may be an important driver of price differences (see Griliches and Mairesse 1998; Fox and Smeets 2011).

lagged prices as instruments. This diminishes the endogeneity concerns, since lagged prices need only to be uncorrelated with the innovation to productivity, η_{jt}. Doraszelski and Jaumandreu (2018) empirically demonstrate that the majority of wage variation in the Spanish manufacturing data set they use is not because of variation in the skill mix of workers and therefore is likely because of geographic and temporal differences in labor markets. This work demonstrates that prices (specifically, lagged prices), when carefully employed, can be a useful source of variation for identification of the production function. However, as also noted in Doraszelski and Jaumandreu (2013), this information is not available in most data sets. Our approach offers an alternative identification strategy that can be employed even when external instruments are not available.

B. *Exploiting First-Order Conditions*

The idea of using first-order conditions for the estimation of production functions dates back to at least the work by Marschak and Andrews (1944), Klein (1953), Solow (1957), and Nerlove (1963),[36] who recognized that, for a Cobb-Douglas production function, there is an explicit relationship between the parameters representing input elasticities and input cost or revenue shares. This observation forms the basis for index number methods (see, e.g., Caves, Christensen, and Diewert 1982) that are used to nonparametrically recover input elasticities and productivity.[37]

More recently, Doraszelski and Jaumandreu (2013, 2018) and Grieco, Li, and Zhang (2016) exploit the first-order conditions for labor and intermediate inputs under the assumption that they are flexibly chosen. Instead of using shares to recover input elasticities, these papers recognize that given a particular parametric form of the production function, the first-order condition for a flexible input (the proxy equation in Levinsohn and Petrin 2003 and Ackerberg, Caves, and Frazer 2015) implies cross-equation parameter restrictions that can be used to aid in identification. Using a Cobb-Douglas production function, Doraszelski and Jaumandreu (2013) show that the first-order condition for a flexible input can be rewritten to replace for productivity in the production function.

[36] Other examples of using first-order conditions to obtain identification include Stone (1954) on consumer demand, Heckman (1974) on labor supply, Hansen and Singleton (1982) on Euler equations and consumption, Paarsch (1992) and Laffont and Vuong (1996) on auctions, and Heckman, Matzkin, and Nesheim (2010) on hedonics.

[37] Index number methods are grounded in three important economic assumptions. First, all inputs are flexible and competitively chosen. Second, the production technology exhibits constant returns to scale, which, while not strictly necessary, is typically assumed to avoid imputing a rental price of capital. Third, and most importantly for our comparison, there are no ex post shocks to output. Allowing for ex post shocks in the index number framework can be relaxed only by assuming that elasticities are constant across firms—i.e., by imposing the parametric structure of Cobb-Douglas.

Combined with observed variation in the prices of labor and intermediate inputs, they are able to estimate the parameters of the production function and productivity.

Doraszelski and Jaumandreu (2018) extend the methodology developed in Doraszelski and Jaumandreu (2013) to estimate productivity when it is non–Hicks neutral, for a CES production function. By exploiting the first-order conditions for both labor and intermediate inputs, they are able to estimate a standard Hicks neutrality and a labor-augmenting component to productivity.

Grieco, Li, and Zhang (2016) also use first-order conditions for both labor and intermediate inputs to recover multiple unobservables. In the presence of unobserved heterogeneous intermediate-input prices, they show that the parametric cross-equation restrictions between the production function and the two first-order conditions, combined with observed wages, can be exploited to estimate the production function and recover the intermediate-input prices. They also show that their approach can be extended to account for the composition of intermediate inputs and the associated (unobserved) component prices.

The paper most closely related to ours is Griliches and Ringstad (1971), which exploits the relationship between the first-order condition for a flexible input and the production function in a Cobb-Douglas parametric setting. They use the average revenue share of the flexible input to measure the output elasticity of flexible inputs. This, combined with the log-linear form of the Cobb-Douglas production function, allows them to then subtract out the term involving flexible inputs. Finally, under the assumption that the nonflexible inputs are predetermined and uncorrelated with productivity (not only with the innovation), they estimate the coefficients for the predetermined inputs.

Our identification solution can be seen as a nonparametric generalization of the Griliches and Ringstad (1971) empirical strategy. Instead of using the Cobb-Douglas restriction, our share equation (11) uses revenue shares to recover input elasticities in a fully nonparametric setting. In addition, rather than subtract out the effect of intermediate inputs from the production function, we instead integrate up the intermediate-input elasticity and take advantage of the nonparametric cross-equation restrictions between the share equation and the production function. Furthermore, we allow for predetermined inputs to be correlated with productivity but uncorrelated with the innovation to productivity.

C. Dynamic Panel

An additional approach employed in the empirical literature on production functions is to use the dynamic panel estimators of Arellano and Bond (1991) and Blundell and Bond (1998, 2000). As discussed in section III, a key insight of the dynamic panel approaches is that by combining panel data observations with some restrictions on the time-series properties of the unobservables, internal instruments can be constructed from within the panel. In contrast to the proxy variable techniques, there is no first stage, and the model consists of a single equation that is an analogue of the proxy variable second stage. Since there is no first stage to recover ε, ω_{jt-1} is solved for from the production function. In the context of our gross output production function described above, we can write

$$y_{jt} = f\left(k_{jt}, l_{jt}, m_{jt}\right) + h\left(y_{jt-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) - \varepsilon_{jt-1}\right) + \eta_{jt} + \varepsilon_{jt}.$$

Notice that the unknown ε_{jt-1} appears inside the nonparametric function h. Typically, these methods proceed under a linearity restriction on h, often an AR(1): $\omega_{jt} = \delta_0 + \delta\omega_{jt-1} + \eta_{jt}$, which implies that[38]

$$\begin{aligned} y_{jt} &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \delta_0 + \delta y_{jt-1} - \delta f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \\ &\quad + \underbrace{\left(-\delta\varepsilon_{jt-1} + \eta_{jt} + \varepsilon_{jt}\right)}_{\psi_{jt}}. \end{aligned} \tag{25}$$

The error ψ_{jt} is then used to construct moment conditions to estimate the model.

In appendix O1, we show that, under the assumptions underlying the proxy variable techniques, an analogue of our identification result in theorem 1 can be obtained for the dynamic panel approaches. As with the proxy variable approach, there are not enough sources of cross-sectional variation available to identify the gross output production function. However, it is important to note that one potential advantage of dynamic panel is that it does not involve inverting for productivity in a first stage. As a result, the scalar unobservability/monotonicity assumption of the proxy literature (assumption 3) is not needed,[39] and dynamic panel methods can accommodate other sources of unobserved variation in the demand for intermediate inputs. This variation could then be used to identify the gross output production function. This would require a version of assumption 5 that includes all inputs in the production function,

[38] Dynamic panel models also typically include fixed effects, which involves additional differencing to remove the fixed effect. For simplicity, we focus here on the case without fixed effects. The essence of our discussion does not depend on whether fixed effects are included. In app. O6, we show that if we impose a linear process for ω, as in the dynamic panel literature, our methodology described in sec. IV can be similarly extended to handle fixed effects by differencing them out.

[39] A related benefit is that these methods do not need to assume anything about $E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}]$.

including intermediate inputs.[40] As pointed out by Ackerberg, Caves, and Frazer (2015), the trade-off is that stronger assumptions are needed on the process for productivity (linearity) and the two components of productivity, ω and ε, cannot be separated.

VII. Empirical Results and Monte Carlo Experiments

In this section, we evaluate the performance of our proposed empirical strategy for estimating the production function and productivity. Using our approach from section V, we estimate a gross output production function using a complete polynomial series of degree two for both the elasticity and the integration constant in the production function. That is, we use

$$\begin{aligned} D_2^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}\right) = {} & \gamma'_0 + \gamma'_k k_{jt} + \gamma'_l l_{jt} + \gamma'_m m_{jt} + \gamma'_{kk} k_{jt}^2 + \gamma'_{ll} l_{jt}^2 \\ & + \gamma'_{mm} m_{jt}^2 + \gamma'_{kl} k_{jt} l_{jt} + \gamma'_{km} k_{jt} m_{jt} + \gamma'_{lm} l_{jt} m_{jt} \end{aligned}$$

to estimate the intermediate-input elasticity,

$$\mathcal{C}_2\left(k_{jt}, l_{jt}\right) = \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_{kk} k_{jt}^2 + \alpha_{ll} l_{jt}^2 + \alpha_{kl} k_{jt} l_{jt}$$

for the constant of integration and

$$h_3\left(\omega_{jt-1}\right) = \delta_0 + \delta_1 \omega_{jt-1} + \delta_2 \omega_{jt-1}^2 + \delta_3 \omega_{jt-1}^3$$

for the Markovian process.

We first illustrate the performance of our approach using Monte Carlo simulations. We then apply our estimator, as well as several extensions of it, to real data using two commonly employed plant-level manufacturing data sets.

A. *Monte Carlo Evidence on Estimator Performance*

Under assumptions 1–5, our procedure generates nonparametric estimates of the production function. To evaluate the performance of our estimator, we first simulate data under these assumptions. To simplify the problem, we abstract away from labor and consider a production function in capital and intermediate inputs only. We begin by examining how our estimator performs under our baseline Monte Carlo specification of

[40] See also Ackerberg, Caves, and Frazer (2015) for a discussion of serially uncorrelated shocks in the context of a value-added production function. Note that not satisfying the proxy variable assumption does not guarantee identification in the presence of a flexible input. For example, unobserved serially correlated intermediate-input price shocks violate the proxy variable assumption. However, this variation generates a measurement problem, since intermediate inputs are typically measured in expenditures (see Grieco, Li, and Zhang 2016).

a Cobb-Douglas production function, using the same basic setup as described in section III.B. To highlight the fact that our approach does not rely on time-series variation in prices, we impose that relative prices are constant over time and thus fix the price of intermediate inputs ρ_t to one. (See app. O4 [Monte Carlo 2] for additional details.) We generate 100 simulated data sets covering 500 firms over 30 periods.

Columns 1 and 2 of table 1 summarize the results of estimating the production function using our nonparametric procedure on 100 simulated data sets. Although the data are generated under a constant output elasticity of intermediate inputs and capital (0.65 and 0.25, respectively), under our nonparametric procedure, the estimated elasticities are allowed to vary across firm and time. Therefore, for each simulation, we calculate three statistics of our estimated elasticities: the mean, the standard deviation, and the fraction that are outside of the (0,1) range. In the table, we report the average of each statistic and its standard error (calculated across the 100 simulations) in parentheses below the point estimates.

As shown in the table, the average mean elasticities of intermediate inputs and capital obtained by our procedure are very close to the true values. This is also true across simulations, as evidenced by the very small standard errors. The standard deviations of the estimated elasticities are also very small, indicating that our procedure is doing a good job of recovering the constant elasticities implied by the Cobb-Douglas specification. Finally, none of the estimated elasticities are either below zero or above one.

While our procedure correctly recovers the lack of variation in elasticities implied by Cobb-Douglas, we also want to evaluate how well our estimator recovers the distribution of elasticities when they are allowed to be heterogeneous across firms and periods in the data. In the remaining columns of table 1, we estimate our model using data generated from both CES and translog production functions,[41] maintaining the other details of our Monte Carlo setup. As with Cobb-Douglas, our procedure does exceptionally well in replicating the true distribution of elasticities for both CES and translog.

The Monte Carlo results summarized in table 1 illustrate that our new identification and estimation strategy performs extremely well under the assumptions described above (assumptions 1–5). Since our approach relies on the first-order condition with respect to a flexible input holding, we also investigate the robustness of our estimator to violations of this assumption. To do so, in appendix O4 (Monte Carlo 3), we discuss

[41] The specific parametrized production functions that we use are $Y_{jt} = (0.25K_{jt}^{0.5} + 0.65M_{jt}^{0.5})^{0.9/0.5} e^{\omega_{jt}+\varepsilon_{jt}}$ for CES and $y_{jt} = 0.25k_{jt} + 0.65m_{jt} + 0.015k_{jt}^2 + 0.015m_{jt}^2 - 0.032k_{jt}m_{jt} + \omega_{jt} + \varepsilon_{jt}$ for translog (in logs).

TABLE 1
Monte Carlo–GNR Estimator Performance

	True Functional Form					
	Cobb-Douglas		CES		Translog	
	At True Parameters (1)	GNR Estimates (2)	At True Parameters (3)	GNR Estimates (4)	At True Parameters (5)	GNR Estimates (6)
Intermediates:						
Average mean elasticity	.6500	.6502	.6747	.6746	.6574	.6572
	. . .	(.0015)	(.0027)	(.0030)	(.0007)	(.0014)
Average standard deviation	0	.0038	.1197	.1193	.0321	.0324
	. . .	(.0012)	(.0014)	(.0018)	(.0004)	(.0012)
Average fraction outside of (0,1)	0	.0000	0	.0000	.0000	.0000
	. . .	(.0000)	. . .	(.0000)	(.0000)	(.0000)
Capital:						
Average mean elasticity	.2500	.2504	.2253	.2198	.2263	.2263
	. . .	(.0065)	(.0027)	(.0071)	(.0006)	(.0082)
Average standard deviation	0	.0091	.1197	.1210	.0333	.0349
	. . .	(.0043)	(.0014)	(.0022)	(.0004)	(.0022)
Average fraction outside of (0,1)	0	.0000	0	.0096	.0000	.0000
	. . .	(.0000)	. . .	(.0055)	(.0000)	(.0000)
Sum:						
Average mean elasticity	.9000	.9006	.9000	.8943	.8836	.8835
	. . .	(.0065)	. . .	(.0074)	(.0002)	(.0080)
Average standard deviation	0	.0093	0	.0227	.0056	.0108
	. . .	(.0044)	. . .	(.0037)	(.0001)	(.0062)

Note.—In this table, we compare estimates of the production function elasticities using our nonparametric procedure with estimates using the true values. We simulate data from thee different parametric production functions: Cobb-Douglas, CES, and translog. See app. O4 for the details. For each parametric form of the production function, the numbers in the first column are computed using the true parameter values. The numbers in the second column are estimated using a complete polynomial series of degree two for each of the two nonparametric functions ($\mathcal{D}$ and $\mathcal{C}$) of our approach. For each simulated data set, we calculate the mean and standard deviation of the output elasticities of capital and intermediate inputs (as well as the sum) across firms and time periods. We also calculate the fraction of the elasticities outside of the range of (0,1). We report the average of these three statistics across each of the simulated data sets, as well as the corresponding standard error (calculated across the 100 simulations). Monte Carlo standard errors are computed by calculating the standard deviation of the statistic of interest across the 100 Monte Carlo samples and are reported in parentheses below the point estimates. Cells with ellipses indicate cases in which there is no variation in a statistic across simulations under the true parameter values. For example, under Cobb-Douglas, the true production function elasticities are constant across simulations. Also, for cases in which a given statistic is identically equal to zero under the true parameter values, we report this as zero with no decimals. For example, under CES, the elasticities are always strictly positive and less than one given our chosen parameter values.

results from a Monte Carlo experiment in which we introduce adjustment frictions in the flexible input into the data-generating process. Specifically, intermediate inputs are now subject to quadratic adjustment costs, ranging from zero adjustment costs to very large adjustment costs. For the largest value of adjustment costs, this would imply that firms in our Chilean and Colombian data sets on average pay substantial adjustment costs for intermediate inputs of almost 10% of the value of total gross output.

We again generate 100 Monte Carlo samples covering 500 firms over 30 periods, and we do this for each of the nine values of adjustment costs. For each sample, we estimate the average capital and intermediate-input elasticities in two ways. As a benchmark, we first obtain estimates using a simple version of dynamic panel with no fixed effects, as described in equation (25). Under dynamic panel, the presence of adjustment costs generates cross-sectional variation in intermediate-input demand (via lagged intermediate inputs) that can be used to identify the model. We then compare these estimates with those obtained via our nonparametric procedure, which assumes adjustment costs of zero. We impose the (true) Cobb-Douglas and AR(1) parametric forms in the estimation of dynamic panel (but not in our nonparametric procedure) to give dynamic panel the best possible chance of recovering the true parameters and to minimize the associated standard errors. We use a constant and k_{jt}, k_{jt-1}, m_{jt-1} as the instruments.

Since the novel part of our procedure relates to the intermediate-input elasticity via the first stage, we focus on the intermediate-input elasticity estimates. The comparison for the capital elasticities is very similar. The results are presented graphically in figures O4.2 and O4.3. As expected, for zero adjustment costs, our procedure recovers the true elasticity very precisely and dynamic panel breaks down. As we increase the level of adjustment costs, the performance of dynamic panel improves, also as expected. Somewhat surprisingly, however, our procedure continues to perform remarkably well, even for the largest values of adjustment costs, with our estimates reflecting only a small bias in the elasticities.

B. *Estimation Results on Chilean and Colombian Data*

Having established that our estimator performs well in Monte Carlo simulations, we now evaluate the performance of our estimator on real data. The first data set we use comes from the Colombian manufacturing census, covering all manufacturing plants with more than 10 employees from 1981 to 1991. This data set has been used in several studies, including Roberts and Tybout (1997); Clerides, Lach, and Tybout (1998); and Das, Roberts, and Tybout (2007). The second data set comes from the census of Chilean manufacturing plants conducted by Chile's Instituto

Nacional de Estadística. It covers all firms from 1979 to 1996 with more than 10 employees. This data set has also been used extensively in previous studies, both in the production function estimation literature (Levinsohn and Petrin 2003) and in the international trade literature (Pavcnik 2002; Alvarez and López 2005).[42]

We estimate separate production functions for the five largest three-digit manufacturing industries in both Colombia and Chile, which are food products (311), textiles (321), apparel (322), wood products (331), and fabricated metal products (381). We also estimate an aggregate specification grouping all manufacturing together.[43] As described above, we use a complete polynomial series of degree two for both the elasticity and the integration constant in the production function and a polynomial of degree three for the Markovian process.[44]

In table 2, for each country-industry pair, we report estimates of the average output elasticities for each input, as well as the sum. We also report the ratio of the average capital and labor elasticities, which measures the capital intensity (relative to labor) of the production technology in each industry. The table includes both estimates from our procedure (labeled "GNR") and, for comparison, estimates obtained from applying simple linear regression (labeled "OLS").

Our estimation approach generates output elasticities that are quite reasonable and precisely estimated, as evidenced by the low standard errors. Intermediate inputs have the highest elasticity, with an average ranging from 0.50 to 0.67 across country/industry. The ranges for labor and capital are 0.22–0.52 and 0.04–0.16, respectively. The sum of the elasticities, a measure of the local returns to scale, is also sensible, ranging from 0.99 to 1.15. Food products (311) and textiles (321) are the most capital-intensive industries in Colombia, and in Chile the most capital-intensive industries are food products, textiles, and fabricated metals (381). In both countries, apparel (322) and wood products (331) are the least capital-intensive industries, even compared with the aggregate specification denoted "All" in the tables.

[42] We construct the variables by adopting the convention used by Greenstreet (2007) with the Chilean data set and employ the same approach with the Colombian data set. In particular, real gross output is measured as deflated revenues. Intermediate inputs are formed as the sum of expenditures on raw materials, energy (fuels plus electricity), and services. Labor input is measured as a weighted sum of blue- and white-collar workers, where blue-collar workers are weighted by the ratio of the average blue-collar wage to the average white-collar wage. Capital is constructed using the perpetual inventory method, where investment in new capital is combined with deflated capital from period $t-1$ to form capital in period t. Deflators for Colombia are obtained from Pombo (1999), and deflators for Chile are obtained from Bergoeing, Hernando, and Repetto (2003).

[43] For all of the estimates that we present, we obtain standard errors by using the nonparametric bootstrap with 200 replications.

[44] We also experimented with higher-order polynomials, and the results were very similar. In a few industries (specifically, those with the smallest number of observations), the results are slightly more heterogeneous, as expected.

TABLE 2

Average Input Elasticities of Output

(Structural vs. Uncorrected OLS Estimates)

	Industry (ISIC Code)											
	Food Products (311)		Textiles (321)		Apparel (322)		Wood Products (331)		Fabricated Metals (381)		All	
	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS
Colombia:												
Labor	.22	.15	.32	.21	.42	.32	.44	.32	.43	.29	.35	.26
	(.02)	(.01)	(.03)	(.02)	(.02)	(.01)	(.05)	(.03)	(.02)	(.02)	(.01)	(.01)
Capital	.12	.04	.16	.06	.05	.01	.04	.03	.10	.03	.14	.06
	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.01)	(.00)
Intermediates	.67	.82	.54	.76	.52	.68	.51	.65	.53	.73	.54	.72
	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.00)	(.00)
Sum	1.01	1.01	1.01	1.03	.99	1.01	.99	1.00	1.06	1.05	1.04	1.04
	(.01)	(.01)	(.02)	(.01)	(.01)	(.01)	(.04)	(.02)	(.01)	(.01)	(.00)	(.00)
Mean (capital)/mean (labor)	.55	.27	.49	.27	.12	.04	.08	.08	.23	.11	.40	.23
	(.08)	(.07)	(.09)	(.06)	(.04)	(.02)	(.05)	(.05)	(.04)	(.04)	(.03)	(.01)
Chile:												
Labor	.28	.17	.45	.26	.45	.29	.40	.20	.52	.32	.38	.20
	(.01)	(.01)	(.03)	(.02)	(.02)	(.02)	(.02)	(.01)	(.03)	(.02)	(.01)	(.01)
Capital	.11	.05	.11	.06	.06	.03	.07	.02	.13	.07	.16	.09
	(.01)	(.00)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.00)	(.00)
Intermediates	.67	.83	.54	.75	.56	.74	.59	.81	.50	.71	.55	.77
	(.00)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.01)	(.00)	(.00)
Sum	1.05	1.05	1.10	1.06	1.08	1.06	1.06	1.04	1.15	1.10	1.09	1.06
	(.01)	(.00)	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)	(.01)	(.01)	(.00)
Mean (capital)/mean (labor)	.39	.28	.24	.22	.14	.12	.18	.12	.25	.21	.43	.42
	(.03)	(.03)	(.04)	(.04)	(.03)	(.03)	(.03)	(.05)	(.03)	(.04)	(.02)	(.02)

Note.—Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates. For each industry, the numbers in the first column are based on a gross output specification using a complete polynomial series of degree two for each of the two nonparametric functions (D and $\mathcal{C}$) of our approach (GNR). The numbers in the second column are also based on a gross output specification and are estimated using a complete polynomial series of degree two with OLS. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry. The "Sum" row reports the sum of the average labor, capital, and intermediate-input elasticities, and the "Mean (capital)/mean (labor)" row reports the ratio of the average capital elasticity to the average labor elasticity.

Our nonparametric procedure also generates distributions of the elasticities across firms that are well-behaved. For any given industry, at most 2% of the labor and intermediate-input elasticities are outside of the range (0,1). For capital, the elasticities are closer to zero on average, but even in the worst case less than 9.4% have values below zero. Not surprisingly, these percentages are highest among the industries with the smallest number of observations.

To evaluate the importance of transmission bias, we compare estimates from our procedure with those using simple linear regression (OLS). A well-known result is that failing to control for transmission bias leads to overestimates of the coefficients on more flexible inputs. The intuition is that the more flexible the input is, the more it responds to productivity shocks and the higher the degree of correlation there is between that input and unobserved productivity. The estimates in table 2 show that the OLS results substantially overestimate the output elasticity of intermediate inputs in every case. The average difference is 34%, which illustrates the importance of controlling for the endogeneity generated by the correlation between input decisions and productivity. The output elasticities of capital and labor are also affected, with OLS underestimating both elasticities. The effect is larger for labor, and as a result, the average elasticity of capital relative to labor is underestimated as well, implying much different factor intensities in the technology. In summary, we find that our approach provides reasonable estimates of the gross output production function while simultaneously correcting for transmission bias.

Given estimates of the production function, we now examine the resulting estimates of productivity. Following Olley and Pakes (1996), we define productivity (in levels) as the sum of the persistent and unanticipated components: $e^{\omega+\varepsilon}$.[45] In table 3, we report estimates of several frequently analyzed statistics of the resulting productivity distributions. In the first three rows for each country, we report ratios of percentiles of the productivity distribution, a commonly used measure of productivity dispersion. As the table illustrates, OLS implies different patterns of productivity heterogeneity. For both countries, the OLS estimates of productivity dispersion are systematically smaller compared with our estimates. As an example, for the case in which we group all industries together (labeled “All” in the table), the 95/5 ratio of productivity is 21% larger for Colombia under our estimates compared with OLS and 16% larger for Chile. The OLS estimates also imply smaller levels of persistence in productivity over time. The average correlation coefficient between current and lagged productivity is 0.64 for our estimates and 0.53 under OLS.

[45] We conduct our analysis using productivity in levels. An alternative would be to use logs. While measuring productivity in levels can exacerbate extreme values, the log transformation is only a good approximation for measuring percentage differences in productivity across groups when these differences are small, which they are not in our data.

The OLS estimates also tend to underestimate the relationship between productivity and other plant characteristics.[46] For example, in almost every industry, we find no evidence of a difference in productivity between exporters and nonexporters under the OLS estimates. After correcting for transmission bias, we find that in many cases exporters are more productive. Examining importers of intermediate inputs, we find an even larger disparity. On average, OLS estimates productivity differences of 1% for Colombia and 6% for Chile. Our estimates imply much larger importer premia of 8% and 13%, respectively. Finally, when we compare firms on the basis of advertising expenditures, not only are there sizable differences in average productivities between OLS and our estimates, but in many cases the sign of the relationship actually changes. When comparing productivity between plants that pay wages above versus below the industry median, the OLS estimates of the differences in productivity are between 28% and 44% smaller for Colombia and between 19% and 44% smaller for Chile.

C. *Robustness Checks and Extensions*

1. Model Fit

In this section, we evaluate the performance/fit of our model. Recall that our model implies that two conditional moment restrictions hold: $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}] = 0$ and $E[\eta_{jt} \mid \mathcal{I}_{jt-1}] = 0$. For estimation, we choose a subset of these restrictions: $E[\varepsilon_{jt} \mid \mathcal{I}_{jt}^{s^\varepsilon}] = 0$ and $E[\eta_{jt} \mid \mathcal{I}_{jt-1}^{s^\eta}] = 0$, where $\mathcal{I}_{jt}^{s^\varepsilon} \subset \mathcal{I}_{jt}$ and $\mathcal{I}_{jt-1}^{s^\eta} \subset \mathcal{I}_{jt-1}$, such that the model is just identified. We do this because our statistical inference is based on results that are valid only for M-estimators. Since we cannot overidentify the model at estimation time, we instead test whether additional restrictions implied by our model are satisfied by our estimates from the just-identified model. A detailed description of our procedure and the corresponding results can be found in appendix O8.

For the individual industry specifications, we cannot reject the hypothesis that the model fits these additional moment restrictions. The lone exception is for fabricated metals (381) in Chile, in which there seems to be some evidence that these additional restrictions are not satisfied. In the specifications that group all industries together, we find that the test rejects the null hypothesis that the additional moments hold. This is not surprising, given the strong restriction imposed by this specification that

[46] As discussed by De Loecker (2013) and Doraszelski and Jaumandreu (2013), one should be careful in interpreting regressions of productivity on characteristics of the firms, to the extent that these characteristics (e.g., exporting or R&D) affect the evolution of productivity and are not explicitly included in the estimation procedure. Our estimates are intended only as a means of comparing OLS with our approach.

TABLE 3
HETEROGENEITY IN PRODUCTIVITY
(Structural vs. Uncorrected OLS Estimates)

	Industry (ISIC Code)											
	Food Products (311)		Textiles (321)		Apparel (322)		Wood Products (331)		Fabricated Metals (381)		All	
	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS	GNR	OLS
Colombia:												
75/25 ratio	1.33	1.16	1.35	1.21	1.29	1.17	1.30	1.23	1.31	1.23	1.37	1.24
	(.02)	(.01)	(.03)	(.01)	(.01)	(.01)	(.04)	(.02)	(.02)	(.01)	(.01)	(.00)
90/10 ratio	1.77	1.42	1.83	1.51	1.66	1.44	1.80	1.57	1.74	1.53	1.86	1.58
	(.05)	(.02)	(.07)	(.04)	(.03)	(.02)	(.12)	(.06)	(.03)	(.02)	(.02)	(.01)
95/5 ratio	2.24	1.74	2.38	1.82	2.02	1.74	2.24	2.01	2.16	1.82	2.36	1.94
	(.08)	(.05)	(.14)	(.08)	(.05)	(.04)	(.22)	(.15)	(.06)	(.04)	(.03)	(.02)
Exporter	.14	.09	.02	−.01	.05	.00	.15	.10	.08	.03	.06	.01
	(.05)	(.04)	(.03)	(.01)	(.03)	(.01)	(.14)	(.09)	(.03)	(.02)	(.01)	(.01)
Importer	.04	−.02	.05	.00	.12	.02	.05	−.03	.10	.05	.11	.04
	(.02)	(.01)	(.04)	(.01)	(.03)	(.01)	(.08)	(.02)	(.02)	(.01)	(.01)	(.01)
Advertiser	−.03	−.07	.08	−.04	.05	−.03	.04	−.02	.05	.00	.03	−.02
	(.02)	(.02)	(.03)	(.02)	(.02)	(.01)	(.04)	(.03)	(.02)	(.01)	(.01)	(.01)
Wages >median	.09	.06	.18	.10	.18	.13	.15	.11	.22	.13	.20	.13
	(.02)	(.02)	(.03)	(.02)	(.02)	(.01)	(.04)	(.03)	(.02)	(.01)	(.01)	(.01)

Chile:												
75/25 ratio	1.37	1.30	1.48	1.40	1.43	1.36	1.50	1.39	1.53	1.46	1.55	1.45
	(.01)	(.00)	(.02)	(.01)	(.02)	(.01)	(.02)	(.01)	(.02)	(.01)	(.01)	(.00)
90/10 ratio	1.90	1.72	2.16	1.97	2.11	1.91	2.32	2.03	2.33	2.14	2.39	2.14
	(.02)	(.01)	(.05)	(.04)	(.05)	(.03)	(.05)	(.04)	(.05)	(.04)	(.02)	(.01)
95/5 ratio	2.48	2.15	2.91	2.57	2.77	2.45	3.11	2.77	3.13	2.80	3.31	2.86
	(.05)	(.02)	(.09)	(.07)	(.09)	(.05)	(.11)	(.07)	(.10)	(.06)	(.04)	(.03)
Exporter	.02	−.01	.02	−.02	.09	.01	.00	−.02	−.01	.00	.03	−.01
	(.02)	(.02)	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)	(.03)	(.02)	(.01)	(.01)
Importer	.14	.03	.10	.04	.14	.06	.15	.07	.11	.06	.15	.09
	(.02)	(.01)	(.02)	(.02)	(.02)	(.01)	(.03)	(.03)	(.02)	(.02)	(.01)	(.01)
Advertiser	.04	.00	.04	.01	.06	.02	.03	.01	.01	.01	.06	.04
	(.01)	(.01)	(.02)	(.01)	(.02)	(.01)	(.01)	(.01)	(.02)	(.02)	(.01)	(.01)
Wages >median	.21	.12	.19	.15	.22	.16	.21	.13	.22	.16	.30	.24
	(.01)	(.01)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.02)	(.01)	(.01)

Note.—Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates. For each industry, the numbers in the first column are based on a gross output specification using a complete polynomial series of degree two for each of the two nonparametric functions (D and $\mathcal{C}$) of our approach (GNR). The numbers in the second column are also based on a gross output specification and are estimated using a complete polynomial series of degree two with OLS. In the first three rows, we report ratios of productivity for plants at various percentiles of the productivity distribution. In the remaining four rows, we report estimates of the productivity differences between plants (as a fraction) on the basis of whether they have exported some of their output, imported intermediate inputs, spent money on advertising, and paid wages above the industry median. For example, in industry 311 for Chile, our estimates imply that a firm that advertises is on average 4% more productive than a firm that does not advertise.

all firms, regardless of industry, use the same production technology. Overall, this evidence suggests that our model fits the data well.

2. Alternative Flexible Inputs

Our approach exploits the first-order condition with respect to a flexible input. We have used intermediate inputs (the sum of raw materials, energy, and services) as the flexible input, as they have commonly been assumed to be flexible in the literature.[47] We believe that this is a reasonable assumption because (*a*) the model period is typically 1 year and (*b*) what is required is that they can be adjusted flexibly at the margin. To the extent that spot markets for commodities exist, including energy and certain raw materials, this enables firms to make such adjustments.[48] However, it may be the case that in some applications researchers do not want to assume that all intermediate inputs are flexible, or they may want to test the sensitivity of their estimates to this assumption.

As a robustness check on our results, we estimate two different specifications of our model in which we allow some of the components of intermediate inputs to be nonflexible. In particular, the production function we estimate is of the form $F(k_{jt}, l_{jt}, rm_{jt}, ns_{jt})e^{\omega_{jt}+\varepsilon_{jt}}$, where rm denotes raw materials and ns denotes energy plus services. In one specification, we assume rm to be nonflexible and ns to be flexible, and in the other specification we assume the opposite. See tables O7.1–O7.4 (tables O7.1–O7.10 and O8.1 are available online) for these results. Overall, the results are sensible, and the comparison to OLS is similar to our main results.

3. Fixed Effects

As we detail in appendix O6, our identification and estimation strategy can be easily extended to incorporate fixed effects in the production function.[49] The production function allowing for fixed effects, a_j, can be written as $Y_{jt} = F(k_{jt}, l_{jt}, m_{jt})e^{a_j+\omega_{jt}+\varepsilon_{jt}}$.[50] A common drawback of models with

[47] For example, see Pavcnik (2002); Kasahara and Rodrigue (2008); Coşar, Guner, and Tybout (2016); and Garcia-Marin and Voigtländer (2019) for recent papers using data from either Chile or Colombia (or both), and for various other countries, see Cooper and Haltiwanger (2006; United States); Amiti and Konings (2007; Indonesia); Aw, Roberts, and Xu (2011; Taiwan); De Loecker (2011; Belgium); and Doraszelski and Jaumandreu (2013; Spain).

[48] Using the same Chilean manufacturing data that we use here, Petrin and Sivadasan (2013) find that the gap between the marginal product and the marginal cost of intermediate inputs is negligible but find large gaps for labor. This is consistent with our assumption that intermediate inputs are flexible but labor is not.

[49] We follow the dynamic panel literature in this case and assume that the process for ω is an AR(1).

[50] See Kasahara, Schrimpf, and Suzuki (2015) for an important extension of our approach to the general case of firm-specific production functions.

fixed effects is that the differencing of the data needed to subtract out the fixed effects can remove a large portion of the identifying information in the data. In the context of production functions, this often leads to estimates of the capital coefficient and returns to scale that are unrealistically low, as well as large standard errors (see Griliches and Mairesse 1998).

In tables O7.5 and O7.6, we report estimates corresponding to those in tables 2 and 3, using our method to estimate the gross output production function allowing for fixed effects. The elasticity estimates for intermediate inputs are exactly the same as in the specification without fixed effects, as the first stage of our approach does not depend on the presence of fixed effects. We do find some evidence in Colombia of the problems mentioned above, as the sample sizes are smaller than those for Chile. Despite this, the estimates are very similar to those from the main specification for both countries, and the larger differences are associated with larger standard errors.

4. Extra Unobservables

As we show in appendix O6, our approach can also be extended to incorporate additional unobservables driving the intermediate-input demand. Specifically, we allow for an additional serially uncorrelated unobservable in the share equation for the flexible input (e.g., optimization error). This introduces some changes to the identification and estimation procedure, but the core ideas are unchanged. In tables O7.7 and O7.8, we report estimates from this alternative specification. Our results are remarkably robust. The standard errors increase slightly, which is not surprising given that we have introduced an additional unobservable into the model. The point estimates, however, are very similar.

5. Relaxing Independence of the Ex Post Shock

To investigate the importance of our assumption that $\mathcal{E} = E[e^{\varepsilon_{jt}}]$ is a constant, in appendix O7 we present estimates in which we allow $E[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}]$ to vary with $\mathcal{I}_{jt}$. In particular, we let it depend on k_{jt} and l_{jt} and regress $e^{\varepsilon_{jt}}$ on (k_{jt}, l_{jt}) to form the expectation. There is some evidence that the expectation varies with these variables (according to the F-test), although the overall explanatory power is quite low, with R^2 values around 1%. As shown in tables O7.9 and O7.10, the results are overall very similar to our baseline estimates in tables 2 and 3.

VIII. Conclusion

In this paper, we show new results regarding the nonparametric identification of gross output production functions in the presence of both flexible and nonflexible inputs under the model structure of the proxy variable approach. We first show that with panel data on output and inputs alone, there are not enough sources of cross-sectional variation for the gross output production function to be identified nonparametrically using either the proxy variable or dynamic panel techniques. We then show that while in theory aggregate price variation can be used to resolve this, Monte Carlo evidence suggests that it may perform poorly in practice.

We offer a new identification strategy (and a simple corresponding estimator) that does not rely on researchers having access to long panels with rich aggregate time-series variation or other sources of exogenous cross-sectional variation. The key to our approach is exploiting the nonparametric cross-equation restrictions between the first-order condition for the flexible inputs and the production function. We also show that our approach can accommodate additional features—for example, fixed effects.

We provide Monte Carlo simulation evidence that our nonparametric procedure performs well in recovering the true underlying production function. Using two commonly employed firm-level production data sets, we show that our nonparametric estimator provides reasonable estimates of the production function elasticities. When we compare our estimates with those obtained by OLS, we find that average output elasticities are biased by at least 23% and as much as 73%. OLS also underestimates the degree of productivity dispersion and the correlation between productivity and other plant characteristics.

As discussed in the introduction, there is a growing interest in the literature in estimating gross output models that include intermediate inputs. The results in this paper should provide researchers with a stronger foundation and additional tools for using gross output production functions in practice.

References

Ackerberg, Daniel A., C. Lanier Benkard, Steven Berry, and Ariel Pakes. 2007. "Econometric Tools for Analyzing Market Outcomes." In *Handbook of Econometrics*, vol. 6, edited by James J. Heckman and Edward E. Leamer, 4171–276. Amsterdam: North-Holland.

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. "Identification Properties of Recent Production Function Estimators." *Econometrica* 83 (6): 2411–51.

Alvarez, Roberto, and Ricardo A. López. 2005. "Exporting and Performance: Evidence from Chilean Plants." *Canadian J. Econ.* 38 (4): 1384–400.

Amiti, Mary, and Jozef Konings. 2007. "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia." *A.E.R.* 97 (5): 1611–38.

Arellano, Manuel, and Stephen Bond. 1991. "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations." *Rev. Econ. Studies* 58 (2): 277–97.

Aw, Bee Yan, Mark J. Roberts, and Daniel Yi Xu. 2011. "R&D Investment, Exporting, and Productivity Dynamics." *A.E.R.* 101 (4): 1312–44.

Baily, Martin N., Charles Hulten, and David Campbell. 1992. “Productivity Dynamics in Manufacturing Plants.” *Brookings Papers Econ. Activity (Microeconomics)* 1992:187–267.

Bartelsman, Eric J., and Mark Doms. 2000. “Understanding Productivity: Lessons from Longitudinal Microdata.” Working Paper no. 2000-19, Board Governors Fed. Reserve System, Washington, DC.

Basu, Susanto, and John G. Fernald. 1997. “Returns to Scale in U.S. Production: Estimates and Implications.” *J.P.E.* 105 (2): 249–83.

Bergoeing, Raphael, Andrés Hernando, and Andrea Repetto. 2003. “Idiosyncratic Productivity Shocks and Plant-Level Heterogeneity.” Working Paper no. 173, Centro Econ. Apl., Univ. Chile.

Bernard, Andrew B., Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum. 2003. “Plants and Productivity in International Trade.” *A.E.R.* 93 (4): 1268–90.

Bernard, Andrew B., and J. Bradford Jensen. 1995. “Exporters, Jobs, and Wages in U.S. Manufacturing: 1976–1987.” *Brookings Papers Econ. Activity (Microeconomics)* 1995:67–119.

———. 1999. “Exceptional Exporter Performance: Cause, Effect or Both?” *J. Internat. Econ.* 47 (1): 1–25.

Bils, Mark, Pete Klenow, and Cian Ruane. 2017. “Misallocation or Mismeasurement?” Working Paper no. 599, Center Internat. Development, Stanford Univ.

Blum, Bernardo S., Sebastian Claro, Ignatius Horstmann, and David A. Rivers. 2019. “The ABCs of Firm Heterogeneity: The Effects of Demand and Cost Differences on Exporting.” Working paper, Univ. Toronto.

Blundell, Richard, and Stephen Bond. 1998. “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models.” *J. Econometrics* 87 (1): 115–43.

———. 2000. “GMM Estimation with Persistent Panel Data: An Application to Production Functions.” *Econometric Rev.* 19 (3): 321–40.

Bond, Stephen, and Måns Söderbom. 2005. “Adjustment Costs and the Identification of Cobb Douglas Production Functions.” IFS Working Paper no. W05/04, Inst. Fiscal Studies, London.

Bruno, Michael. 1978. “Duality, Intermediate Inputs, and Value-Added.” In *Production Economics: A Dual Approach to Theory and Applications*, vol. 2, edited by Melvyn Fuss and Daniel McFadden, 3–16. Amsterdam: North-Holland.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert. 1982. “The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity.” *Econometrica* 50 (6): 1393–414.

Chen, Xiaohong. 2007. “Large Sample Sieve Estimation of Semi-nonparametric Models.” In *Handbook of Econometrics*, vol. 6, edited by James J. Heckman and Edward E. Leamer, 5549–632. Amsterdam: Elsevier.

Clerides, Sofronis K., Saul Lach, and James R. Tybout. 1998. “Is Learning by Exporting Important? Micro-dynamic Evidence from Colombia, Mexico, and Morocco.” *Q.J.E.* 113 (3): 903–47.

Collard-Wexler, Allan. 2010. “Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry.” Working paper, Dept. Econ., New York Univ. Stern School Bus.

Cooper, Russell W., and John C. Haltiwanger. 2006. “On the Nature of Capital Adjustment Costs.” *Rev. Econ. Studies* 73 (3): 611–33.

Coşar, A. Kerem, Nezih Guner, and James Tybout. 2016. “Firm Dynamics, Job Turnover, and Wage Distributions in an Open Economy.” *A.E.R.* 106 (3): 625–63.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout. 2007. “Market Entry Costs, Producer Heterogeneity, and Export Dynamics.” *Econometrica* 75 (3): 837–73.

De Loecker, Jan. 2011. “Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity.” *Econometrica* 79 (5): 1407–51.

———. 2013. “Detecting Learning by Exporting.” *American Econ. J.: Microeconomics* 5 (3): 1–21.

De Loecker, Jan, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik. 2016. “Prices, Markups, and Trade Reform.” *Econometrica* 84 (2): 445–510.

Dhrymes, Phoebus J. 1991. “The Structure of Production Technology Productivity and Aggregation Effects.” CES Working Paper no. 91-5, Center Econ. Studies, Washington, DC.

Doraszelski, Ulrich, and Jordi Jaumandreu. 2013. “R&D and Productivity: Estimating Endogenous Productivity.” *Rev. Econ. Studies* 80 (4): 1338–83.

———. 2018. “Measuring the Bias of Technological Change.” *J.P.E.* 126 (3): 1027–84.

Foster, Lucia, John Haltiwanger, and Chad Syverson. 2008. “Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?” *A.E.R.* 98 (1): 394–425.

Fox, Jeremy, and Valérie Smeets. 2011. “Does Input Quality Drive Measured Differences in Firm Productivity?” *Internat. Econ. Rev.* 52 (4): 961–89.

Gandhi, Amit, Salvador Navarro, and David A. Rivers. 2017. “How Heterogeneous Is Productivity? A Comparison of Gross Output and Value Added.” CHCP Working Paper no. 201727, Center Human Capital and Productivity, Univ. Western Ontario.

Garcia-Marin, Alvaro, and Nico Voigtländer. 2019. “Exporting and Plant-Level Efficiency Gains: It’s in the Measure.” *J.P.E.* 127 (4): 1777–825.

Goldberger, Arthur S. 1968. “The Interpretation and Estimation of Cobb-Douglas Functions.” *Econometrica* 35 (3/4): 464–72.

Greenstreet, David. 2007. “Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules.” Economics Series Working Paper no. 345, Dept. Econ., Univ. Oxford.

Grieco, Paul, Shengyu Li, and Hongsong Zhang. 2016. “Production Function Estimation with Unobserved Input Price Dispersion.” *Internat. Econ. Rev.* 57 (2): 665–90.

Griliches, Zvi, and Jacques Mairesse. 1998. “Production Functions: The Search for Identification.” In *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, edited by Steinar Strøm, 169–203. New York: Cambridge Univ. Press.

Griliches, Zvi, and Vidar Ringstad. 1971. *Economies of Scale and the Form of the Production Function: An Econometric Study of Norwegian Manufacturing Establishment Data.* Amsterdam: North-Holland.

Hahn, Jinyong, Zhipeng Liao, and Geert Ridder. 2018. “Nonparametric Two-Step Sieve M Estimation and Inference.” *Econometric Theory* 34 (6): 1281–324.

Halpern, László, Miklós Koren, and Adam Szeidl. 2015. “Imported Inputs and Productivity.” *A.E.R.* 105 (12): 3660–703.

Hansen, Lars Peter, and Thomas J. Sargent. 1980. “Formulating and Estimating Dynamic Linear Rational Expectations Models.” *J. Econ. Dynamics and Control* 2:7–46.

Hansen, Lars Peter, and Kenneth J. Singleton. 1982. “Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models.” *Econometrica* 50 (5): 1269–86.

Heckman, James J. 1974. “Shadow Prices, Market Wages, and Labor Supply.” *Econometrica* 42 (4): 679–94.

Heckman, James J., Rosa L. Matzkin, and Lars Nesheim. 2010. “Nonparametric Identification and Estimation of Nonadditive Hedonic Models.” *Econometrica* 78 (5): 1569–91.

Heckman, James J., and Richard Robb. 1985. “Alternative Methods for Evaluating the Impact of Interventions.” In *Longitudinal Analysis of Labor Market Data*, edited by James J. Heckman and Burton Singer, 156–246. New York: Cambridge Univ. Press.

Heckman, James J., and Edward J. Vytlacil. 2007. “Econometric Evaluation of Social Programs. II. Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments.” In *Handbook of Econometrics*, vol. 6, edited by James J. Heckman and Edward E. Leamer, 4875–5143. Amsterdam: North-Holland.

Horowitz, Joel L. 2001. “The Bootstrap.” In *Handbook of Econometrics*, vol. 5, edited by James J. Heckman and Edward E. Leamer, 3159–228. Amsterdam: North-Holland.

Houthakker, Hendrik S. 1950. “Revealed Preference and the Utility Function.” *Economica* 17 (66): 159–74.

Kasahara, Hiroyuki, and Joel Rodrigue. 2008. “Does the Use of Imported Intermediates Increase Productivity? Plant-Level Evidence.” *J. Development Econ.* 87 (1): 106–18.

Kasahara, Hiroyuki, Paul Schrimpf, and Michio Suzuki. 2015. “Identification and Estimation of Production Function with Unobserved Heterogeneity.” Working paper, Vancouver School Econ., Univ. British Columbia.

Klein, Lawrence R. 1953. *A Textbook of Econometrics*. Evanston, IL: Row, Peterson.

Klein, Roger, and Francis Vella. 2010. “Estimating a Class of Triangular Simultaneous Equations Models without Exclusion Restrictions.” *J. Econometrics* 154 (2): 154–64.

Koopmans, Tjalling C., Herman Rubin, and Roy Bergh Leipnik. 1950. “Measuring the Equation Systems of Dynamic Economics.” In *Statistical Inference in Dynamic Economic Models*, vol. 10, edited by Tjalling C. Koopmans, 53–237. New York: Wiley.

Laffont, Jean-Jacques, and Quang Vuong. 1996. “Structural Analysis of Auction Data.” *A.E.R.* 86 (2): 414–20.

Levinsohn, James, and Amil Petrin. 2003. “Estimating Production Functions Using Inputs to Control for Unobservables.” *Rev. Econ. Studies* 70 (2): 317–42.

Lewbel, Arthur. 2012. “Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models.” *J. Bus. and Econ. Statis.* 30 (1): 67–80.

Lucas, Robert E., and Thomas J. Sargent, eds. 1981. *Rational Expectations and Econometric Practice*. Minneapolis: Univ. Minnesota Press.

Marschak, Jacob, and William H. Andrews. 1944. “Random Simultaneous Equations and the Theory of Production.” *Econometrica* 12 (3/4): 143–205.

Nerlove, Marc. 1963. “Returns to Scale in Electricity Supply.” In *Measurement in Economics*, edited by Carl F. Christ, Milton Friedman, Leo A. Goodman, et al., 167–200. Stanford, CA: Stanford Univ. Press.

Newey, Whitney K., and James L. Powell. 2003. “Instrumental Variable Estimation of Nonparametric Models.” *Econometrica* 71 (5): 1565–78.

Newey, Whitney K., James L. Powell, and F. Vella. 1999. “Nonparametric Estimation of Triangular Simultaneous Equations Models.” *Econometrica* 67 (3): 565–603.

Oberfield, Ezra. 2013. “Productivity and Misallocation during a Crisis: Evidence from the Chilean Crisis of 1982.” *Rev. Econ. Dynamics* 16 (1): 100–119.

Olley, G. Steven, and Ariel Pakes. 1996. “The Dynamics of Productivity in the Telecommunications Equipment Industry.” *Econometrica* 64 (6): 1263–97.

Paarsch, Harry J. 1992. “Deciding between the Common and Private Value Paradigms in Empirical Models of Auctions.” *J. Econometrics* 51 (1/2): 191–215.

Pavcnik, Nina. 2002. “Trade Liberalization Exit and Productivity Improvements: Evidence from Chilean Plants.” *Rev. Econ. Studies* 69 (1): 245–76.

Petrin, Amil, and Jagadeesh Sivadasan. 2013. “Estimating Lost Output from Allocative Inefficiency, with an Application to Chile and Firing Costs.” *Rev. Econ. and Statis.* 95 (1): 286–301.

Pombo, Carlos. 1999. “Productividad Industrial en Colombia: Una Aplicacion de Numeros Indicey.” *Rev. Econ. Rosario* 2 (3): 107–39.

Pozzi, Andrea, and Fabiano Schivardi. 2016. “Demand or Productivity: What Determines Firm Growth?” *RAND J. Econ.* 47 (3): 608–30.

Rigobon, Roberto. 2003. “Identification through Heteroskedasticity.” *Rev. Econ. and Statis.* 85 (4): 777–92.

Roberts, Mark J., and James R. Tybout. 1997. “The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs.” *A.E.R.* 87 (4): 545–64.

Sargent, Thomas J. 1978. “Rational Expectations, Econometric Exogeneity, and Consumption.” *J.P.E.* 86 (4): 673–700.

Solow, Robert M. 1957. “Technical Change and the Aggregate Production Function.” *Rev. Econ. and Statis.* 39 (3): 312–20.

Stone, Richard. 1954. “Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand.” *Econ. J.* 64 (255): 511–27.

Syverson, Chad. 2004. “Product Substitutability and Productivity Dispersion.” *Rev. Econ. and Statis.* 86 (2): 534–50.

———. 2011. “What Determines Productivity?” *J. Econ. Literature* 49 (2): 326–65.

Wooldridge, Jeffrey M. 2009. “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables.” *Econ. Letters* 104 (3): 112–14.