2018

2018-1 On the Identification of Gross Output Production Functions

Amit Gandhi

Salvador Navarro

David Rivers

On the Identification of Gross Output Production Functions

by

Amit Gandhi, Salvador Navarro and David Rivers

Working Paper # 2018-1 **May 2018**

Western

Centre for Human Capital and Productivity (CHCP)

Working Paper Series

Department of Economics
Social Science Centre
Western University
London, Ontario, N6A 5C2
Canada

On the Identification of Gross Output Production Functions

Amit Gandhi, Salvador Navarro, David Rivers*

May 17, 2017

Abstract

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.

*We would like to thank Dan Ackerberg, 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, Lance Lochner, Aureo de Paula, Amil Petrin, Mark Roberts, Nicolas Roys, Todd Stinebrickner, Chad Syverson, Chris Taber, Quang Vuong, and especially Tim Conley for helpful discussions. This paper has also benefited from detailed comments by the editor and three anonymous referees. We would also like to 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?". First draft: May 2006. Amit Gandhi is at the University of Wisconsin-Madison, E-mail: agandhi@ssc.wisc.edu. Salvador Navarro is at the University of Western Ontario, E-mail: snavarr@uwo.ca. David Rivers is at the University of Western Ontario, E-mail: drivers2@uwo.ca.

1 Introduction

The identification and estimation of production functions using data on firm inputs and output is among the oldest empirical problems in 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 has become increasingly available for many countries, and which motivates a variety of industry equilibrium models based on patterns of productivity heterogeneity found in this 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) (henceforth GM) concluded that the standard econometric solutions to correct the transmission bias, i.e., using firm fixed effects or instrumental variables, are both theoretically problematic and unsatisfactory in practice (see also Ackerberg et al., 2007 for a more recent review). 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 et al., 1982) recover the production function and productivity by equating each input's output elasticity to its input share. However, when some inputs 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.

[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, Bernard and Jensen, 1999, Bernard et al., 2003), wages (Baily et al., 1992), etc. See Syverson (2011) for a review of this literature.

[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.

Instead, transmission bias is addressed by imposing assumptions on the economic environment, which allow 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; Ackerberg et al., 2015; Wooldridge, 2009, henceforth OP, LP, ACF, and Wooldridge, respectively) that are now prevalent in the applied literature on production function estimation.[3] Most of these papers (with the exception of LP) focus on some form of a value-added production function. Recent work by ACF 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 et al., 2015; and De Loecker et al., 2016). The macroeconomics literature on misallocation is also now employing gross output models of firm-level production (Oberfield, 2013 and Bils et al., 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 et al., 2008; Pozzi and Schivardi, 2016; and Blum et al., 2017). 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 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 non-identified 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 indices 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

[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 OP, LP, ACF, and Wooldridge.

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

(2013) (henceforth DJ) 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 LP and ACF), 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 DJ, 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 and 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 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 et al. (2016) shows, our

[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 the book by Lucas and Sargent (1981) for many examples of the use of cross-equation restrictions in the context of the rational expectations literature.

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, 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 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 to 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 2 we describe the model and set up the firm's problem. In Section 3 we examine the extent to which the proxy variable/dynamic panel methods can be applied to identify the gross output production function. Section 4 presents our nonparametric identification strategy. In Section 5 we describe our estimation strategy. Section 6 compares our approach to the related literature. In Section 7 we present estimates from our procedure applied to Monte Carlo simulated data as well as plant-level data from Colombia and Chile. Section 8 concludes.

2 The Model

In this section we describe the economic model of the firm that we study. We focus attention in the main body 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 Online Appendix O5.

2.1 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 will be denoted by $(Y_{jt}, K_{jt}, L_{jt}, M_{jt})$ respectively, and their log values will be 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, i.e., the data takes 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 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 $x_{jt} \in \mathcal{I}_{jt}$, i.e., the amount of the input employed in period t, is in the firm's information set for that period, then we say 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 the input is *dynamic.* If an input is neither predetermined nor dynamic, then we say it is *flexible*. We refer to inputs that are predetermined, dynamic, or both as *non-flexible.*

2.2 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*

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

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 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, $P_\varepsilon(\varepsilon_{jt} \mid \mathcal{I}_{jt}) = P_\varepsilon(\varepsilon_{jt})$.*

[6] Formally the firm's information set is the sigma-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.

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 $E[\eta_{jt} \mid \mathcal{I}_{jt-1}] = 0$. 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 which 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 that it allows 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 non-flexible inputs. For simplicity, we mainly focus on the case of a single flexible input in the model (but see Online Appendix O5), 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 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.

[7] It is straightforward to allow the distribution of $P_\omega(\omega_{jt} \mid \mathcal{I}_{jt-1})$ to depend upon 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 DJ 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.

2.3 The Firm's Problem

The proxy variable literature of LP/ACF/Wooldridge uses a flexible input demand, intermediate inputs, to proxy for the unobserved persistent productivity ω.[10] In order 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 et al., 2015). We formalize this in the following assumption.

Assumption 3. *The scalar unobservablility and strict monotonicity assumptions of LP/ACF/Wooldridge 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 strictly monotone in a single unobservable ω_{jt}.

We follow the same setup used by both LP and ACF 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 market, 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

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

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

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

[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] OP does not include intermediate inputs in the model.

[12] See Appendix A in LP and pg. 2429 in ACF.

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 $d_t \equiv \ln\left(\frac{\rho_t}{P_t}\right) - \ln \mathcal{E}$. 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 ($P_\varepsilon\left(\varepsilon_{jt} \mid \mathcal{I}_{jt}\right) = P_\varepsilon\left(\varepsilon_{jt}\right)$). If instead only mean independence of ε_{jt} were assumed ($E\left[\varepsilon_{jt} \mid \mathcal{I}_{jt}\right] = 0$), we would have $P_t \frac{\partial}{\partial M_{jt}} F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}} \mathcal{E}\left(\mathcal{I}_{jt}\right) = \rho_t$, and hence $m_{jt} = \mathbb{M}_t\left(k_{jt}, l_{jt}, \omega_{jt}, \mathcal{I}_{jt}\right)$. Assumption 3 is therefore implicitly imposing that, if $\mathcal{E}\left(\mathcal{I}_{jt}\right)$ 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 and GM). 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 $\frac{\partial}{\partial x_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right)$.

3 The Proxy Variable Framework and Gross Output

Both the dynamic panel literature and the proxy literature of OP/LP/ACF/Wooldridge 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 LP, which employs a gross output specification.

[13] Intermediate inputs, however, may still be used as the proxy variable for productivity (see ACF).

However, previous work by Bond and Söderbom (2005) and ACF has identified an identification problem with the LP procedure. Therefore, in this section we examine whether the modified proxy variable approach developed by ACF 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:

$$E[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, d_t] = f(k_{jt}, l_{jt}, m_{jt}) + \mathbb{M}^{-1}(k_{jt}, l_{jt}, m_{jt}) + d_t \equiv \phi(k_{jt}, l_{jt}, m_{jt}) + d_t, \quad (6)$$

where d_t is a time fixed effect. Clearly no structural elasticities can be identified from this regression (the "first stage"), in particular the flexible input elasticity, $\frac{\partial}{\partial m_{jt}} f(k_{jt}, l_{jt}, m_{jt})$. 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 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 re-write the production function as:

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

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

[14] We restrict our attention in the main body to the use of intermediate inputs as a proxy versus the original proxy variable strategy of OP that uses investment. As LP argued, the fact that investment is often zero in plant-level data leads to practical challenges in using the OP approach, and as a result using intermediate inputs as a proxy has become the preferred strategy in applied work. In Online Appendix O1, we show that our results extend to the case of using investment instead, as well as to the use of dynamic panel methods.

[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 only be formed for observations in which the proxy variable, intermediate input demand (or investment in OP), 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 non-positive values of the proxy, and this probability is a function of the other state variables of the firm.

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 2.1 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]

3.1 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

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

This implies that the only sources of variation left in m_{jt} after conditioning on $(k_{jt}, l_{jt}, d_{t-1}, k_{jt-1}, l_{jt-1}, m_{jt-1}) \in \Gamma_{jt}$ (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.

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

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

The unknown functions underlying equation (9) are given by (f, h), since $\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1}$ 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

[16] The idea that one can use expectations conditional on lagged information sets, in order 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 DJ 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).

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 $E\left[\eta_{jt} + \varepsilon_{jt} \mid \Gamma_{jt}\right] = 0$) to equation (7) is insufficient to identify 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*

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

for any $a \in (0,1)$, that satisfy 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 $E\left[y_{jt} \mid \Gamma_{jt}\right] = E\left[\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t \mid \Gamma_{jt}\right]$. Hence, for any (f, h) that satisfy (9), it must be the case that

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

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

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

[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 $\phi\left(k_{jt}, l_{jt}, m_{jt}\right) = f\left(k_{jt}, l_{jt}, m_{jt}\right)$).

[20] In Online Appendix O1 we show 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.

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

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

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. □

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 Online Appendix 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*

$$\begin{aligned} \tilde{f}\left(k_{jt},l_{jt},m_{jt}\right) &\equiv \left(1-a\right)f^{0}\left(k_{jt},l_{jt},m_{jt}\right)+a\phi\left(k_{jt},l_{jt},m_{jt}\right) \\ \tilde{h}\left(x\right) &\equiv ad_{2}+\left(1-a\right)h^{0}\left(\frac{1}{\left(1-a\right)}\left(x-ad_{1}\right)\right) \end{aligned}$$

for $t = 1, 2$ and for any $a \in (0, 1)$, that satisfy 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. □

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, then*

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

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

or

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

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

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. □

The result in Theorem 1 and its two corollaries is a useful benchmark, as it relates directly 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 instrumental variables 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 model for output, y_{jt}, there is a model for the proxy variable, in this case intermediate inputs, m_{jt}. In Online Appendix O3, we extend our result to the full model involving f, h, and $\mathbb{M}$, using the full distribution of the data.

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

3.1.1 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.

In order to evaluate the performance of using time series variation as a source of identification, we conduct several Monte Carlo experiments. The parameters of the data generating process are chosen to roughly match the properties of our Chilean and Colombian datasets, as well as the variances of our productivity estimates, described below in Section 7. A full description of the setup is provided in the Appendix. The key features are as follows. Firms maximize the expected stream of future discounted profits. 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. Productivity is assumed to evolve according to an AR(1) process with a persistence parameter of 0.8. 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.

We construct 12 different panel structures: 200 vs. 500 firms and 3, 5, 10, 20, 30, and 50 periods. For each panel, we simulate 500 datasets based on four different levels of variation in relative prices. The first two levels of time series variation correspond to what we observe in our Colombian and Chilean datasets, respectively. In addition, we create a version with twice the degree of what we observe for Chile (the largest of the two), and another corresponding to 10 times the observed variation. We estimate a version of the proxy variable technique applied to gross output, as described above, using intermediate inputs as the proxy. In order 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 level of time series variation observed in the data, 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 what we observe, 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 Figure O4.1 in the Online Appendix).

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.[22] 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, 2015), which we discuss in more detail in Section 6. 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.

4 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 and productivity, even absent additional sources of exogenous variation in flexible inputs.[23] 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 (see Klein, 1953; Solow, 1957; and Nerlove, 1963 for early examples, and Doraszelski and Jaumandreu, 2013, 2015 more recently).[24] 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

[22] It may be possible to achieve identification in the absence of exclusion restrictions by imposing additional restrictions (see Koopmans et al., 1950 and Heckman and Robb, 1985). One example is using heteroskedasticity restrictions (see e.g., Rigobon, 2003; Klein and Vella, 2010; and 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.

[23] Please see Online Appendix O5 for the extension to the case of multiple flexible inputs.

[24] See Section 6 for a more detailed discussion of this literature.

such generality is that the proxy variable assumption—Assumption 3—already presumes 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 $s_{jt} \equiv \ln \frac{\rho_t M_{jt}}{P_t Y_{jt}}$ is the (log) intermediate input share of output. In the following theorem we prove that, since $E\left[\varepsilon_{jt} \mid k_{jt}, l_{jt}, m_{jt}\right] = 0$, 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 that $\frac{\rho_t}{P_t}$ (or the relative price-deflator) is observed, the share regression in equation* (11) *nonparametrically identifies the flexible input elasticity $\frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right)$ of the production function almost everywhere in (k_{jt}, l_{jt}, m_{jt}).*

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

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

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

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

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

can be identified.[25] This allows us to identify the flexible input elasticity as

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

□

[25] DJ suggest using this approach to identify this constant in the context of a Cobb-Douglas production function.

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 $P_{\varepsilon}\left(\varepsilon_{jt} \mid \mathcal{I}_{jt}\right) = P_{\varepsilon}\left(\varepsilon_{jt}\right)$ (in Assumption 2). Much of the previous literature assumes only mean independence $E\left[\varepsilon_{jt} \mid \mathcal{I}_{jt}\right] = 0$. 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, $\mathcal{E}\left(\mathcal{I}_{jt}\right) \equiv E\left[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}\right]$. Since ε_{jt} (and hence $e^{\varepsilon_{jt}}$) is identified from the share regression (12), $\mathcal{E}\left(\mathcal{I}_{jt}\right)$ can also be identified. In terms of the proof, the elasticity would then be obtained as $D\left(k_{jt}, l_{jt}, m_{jt}\right) = \frac{D^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}, \mathcal{I}_{jt}\right)}{\mathcal{E}\left(\mathcal{I}_{jt}\right)}$, where notice that now $D^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}, \mathcal{I}_{jt}\right)$ depends on $\mathcal{I}_{jt}$, and hence the share regression would need to be adjusted accordingly.[26]

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.[27] By the fundamental theorem of calculus we have

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

Subtracting equation (15) from the production function, and re-arranging terms we have

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

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 in order to generate moments based on the panel structure of the data and recover $\mathscr{C}\left(k_{jt}, l_{jt}\right)$. By replacing for ω in equation (16), we have

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

Since $\left(\mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}\right)$ are all known to the firm at period $t-1$ and $\left(k_{jt}, l_{jt}\right)$ are predetermined, we

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

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

have the orthogonality $E\left[\eta_{jt} \mid k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}\right] = 0$ which implies

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

A regression of $\mathcal{Y}_{jt}$ on $(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1})$ identifies the LHS 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 $\mathscr{C}$ can be separately identified from h up to an additive constant.[28]

We now establish this result formally in Theorem 3 based on the above discussion. In order 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 et al., 1999).

Assumption 5. *For each point* $\bar{\mathcal{Y}}_{jt}, \bar{k}_{jt-1}, \bar{l}_{jt-1}$ *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* $\bar{\mathcal{Y}}_{jt}, \bar{k}_{jt-1}, \bar{l}_{jt-1}$ *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 29, in the proof of Theorem 3 below, 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* $\frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right)$ *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 1 for any (k_{jt}, l_{jt}, m_{jt}) in the support of the data there is a set

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

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

also contained in the support for some $\underline{\mathrm{m}}(k_{jt}, l_{jt})$. Hence with probability 1 the integral

$$\int_{\underline{\mathrm{m}}(k_{jt},l_{jt})}^{m_{jt}} \frac{\partial}{\partial m_{jt}} f(k_{jt}, l_{jt}, m)\, dm = f(k_{jt}, l_{jt}, m_{jt}) + \mathscr{C}(k_{jt}, l_{jt})$$

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 $\frac{\partial}{\partial m_{jt}} f(k_{jt}, l_{jt}, m_{jt})$ over the support of the data, then they can only differ by an additive function $\mathscr{C}(k_{jt}, l_{jt})$. 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

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

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) $\mu(k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1}) = -\mathscr{C}(k_{jt}, l_{jt}) + h(\mathcal{Y}_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1}))$. Let $\tilde{\mathscr{C}}, \tilde{h}$ be an alternative pair of functions. $(\mathscr{C}, h)$ and $\tilde{\mathscr{C}}, \tilde{h}$ are observationally equivalent if and only if

$$-\mathscr{C}(k_{jt}, l_{jt}) + h(\mathcal{Y}_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1})) = -\tilde{\mathscr{C}}(k_{jt}, l_{jt}) + \tilde{h}\ \mathcal{Y}_{jt-1} + \tilde{\mathscr{C}}(k_{jt-1}, l_{jt-1}) \quad , \tag{20}$$

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}

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

for $z \in \{k_{jt}, l_{jt}\}$ and for all x_{jt} in its support, which implies $\mathscr{C}(k_{jt}, l_{jt}) - \tilde{\mathscr{C}}(k_{jt}, l_{jt}) = c$ for a constant c for almost all x_{jt}.[29] Thus we have shown the production function is identified up to a constant. □

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

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

$$\mathscr{C}\ k^*_{jt}, l^*_{jt} = -E\ \mathcal{Y}_{jt} \mid k^*_{jt}, l^*_{jt}, \bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1} + E\ \mathcal{Y}_{jt} \mid k'_{jt}, l'_{jt}, \bar{\mathcal{Y}}_{jt-1}, \bar{k}_{jt-1}, \bar{l}_{jt-1} + \mathscr{C}\ k'_{jt}, l'_{jt} \quad ,$$

again up to the constant c.

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 like 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) based on dynamic panel methods alone. We revisit this in our discussion of dynamic panel methods in Section 6.3.

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 $t \in \{2, ..., T\}$.

5 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 $\mathscr{C}$ 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, $\sum_{jt} \varepsilon_{jt}^2$, as our objective function. For example, for a complete polynomial of degree two, our estimator would solve:

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

The solution to this problem is an estimator

$$D_r^{\mathcal{E}}(k_{jt}, l_{jt}, m_{jt}) = \sum_{r_k + r_l + r_m \leq r} \gamma'_{r_k, r_l, r_m} k_{jt}^{r_k} l_{jt}^{r_l} m_{jt}^{r_m}, \text{ with } r_k, r_l, r_m \geq 0 \tag{21}$$

of the elasticity up to the constant $\mathcal{E}$, as well as the residual ε_{jt} corresponding to the ex-post shocks to production.[30] Since we can estimate $\hat{\mathcal{E}} = \frac{1}{JT} \sum_{j,t} e^{\hat{\varepsilon}_{jt}}$, we can recover $\hat{\gamma} \equiv \frac{\hat{\gamma}}{\hat{\mathcal{E}}}$, and thus estimate $\frac{\partial}{\partial m_{jt}} f(k_{jt}, l_{jt}, m_{jt})$ 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:

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

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

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

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

In the second step, in order to recover the constant of integration $\mathscr{C}$ 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 $\mathscr{C}(k_{jt}, l_{jt})$ to contain no constant. That is, we use

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

and

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

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:

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

[30] 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 like the one we propose. See Chen (2007).

We can then use moments of the form $E\left[\eta_{jt}k_{jt}^{\tau_k}l_{jt}^{\tau_l}\right]=0$ and $E\left[\eta_{jt}\hat{\mathcal{Y}}_{jt-1}^{a}\right]=0$ to form a standard sieve moment criterion function to estimate (α,δ).[31] 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 are the nonlinear least squares moments corresponding to the share equation.

Under the just-identified case described above, our two-step sieve procedure is a sieve-M estimator. Therefore we can apply the numerical equivalence results of Hahn et al. (2016) and conduct inference as if our sieve was the true parametric structure.[32] In order to compute standard errors for the functionals of interest (e.g., elasticities), we employ a nonparametric bootstrap (see e.g., Horowitz, 2001).[33]

6 Relationship to Literature

6.1 Price Variation as an Instrument

Recall that Theorem 1 (and its extensions in the Online Appendix) 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 3, 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 GM and Ackerberg et al., 2007). First, in many firm-level production datasets, firm-specific prices

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

[32] One could also use higher-order moments, as well as lags of inputs, to estimate an over-identified 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.

[33] In Online Appendix O4, we present Monte Carlo simulations which show that our bootstrap procedure has the correct coverage for the nonparametric estimates.

are simply not observed. Second, even if price variation is observed, in order 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.[34]

This is not to say that if one can isolate exogenous price variation (for example if prices vary due to segmented geographic markets or due to policy shocks), it cannot be used to aid in identification. The point is that just 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 DJ, is to use lagged prices as instruments. This diminishes the endogeneity concerns, since lagged prices only need to be uncorrelated with the innovation to productivity, η_{jt}. Doraszelski and Jaumandreu (2015) demonstrate empirically that the majority of wage variation in the Spanish manufacturing dataset they use is not due to variation in the skill mix of workers, and therefore is likely due to 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 DJ, this information is not available in most datasets. Our approach offers an alternative identification strategy that can be employed even when external instruments are not available.

6.2 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)[35] 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 et al., 1982) that are used to nonparametrically recover input elasticities and productivity.[36]

[34]Recent work has suggested that quality differences may be an important driver of price differences (see GM and Fox and Smeets, 2011).

[35]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 et al. (2010) on hedonics.

[36]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 in order to avoid imputing a rental price of capital. Third, and most importantly for our

More recently, Doraszelski and Jaumandreu (2013, 2015) and Grieco et al. (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 LP/ACF) implies cross-equation parameter restrictions that can be used to aid in identification. Using a Cobb-Douglas production function, DJ show that the first-order condition for a flexible input can be re-written to replace for productivity in the production function. 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 (2015) extend the methodology developed in DJ 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 neutral and a labor-augmenting component to productivity.

Grieco et al. (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 non-flexible inputs are predetermined and uncorrelated with productivity (not just 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,

comparison, there are no ex-post shocks to output. Allowing for ex-post shocks in the index number framework can only be relaxed by assuming that elasticities are constant across firms, i.e., by imposing the parametric structure of Cobb-Douglas.

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 just the innovation to productivity.

6.3 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 3, 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[37]

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

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

In Online 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

[37] 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 or not fixed effects are included. In Online Appendix O5, we show that if we impose a linear process for ω, as in the dynamic panel literature, our methodology described in Section 4 can be similarly extended to handle fixed effects by differencing them out.

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,[38] 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, including intermediate inputs.[39,40] As pointed out by ACF, 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.

7 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 5, we estimate a gross output production function using a complete polynomial series of degree 2 for both the elasticity and the integration constant in the production function. That is, we use

$$\begin{aligned} D_2^{\mathcal{E}}\left(k_{jt}, l_{jt}, m_{jt}\right) &= \gamma_0 + \gamma_k k_{jt} + \gamma_l l_{jt} + \gamma_m m_{jt} + \gamma_{kk} k_{jt}^2 + \gamma_{ll} l_{jt}^2 \\ &\quad + \gamma_{mm} m_{jt}^2 + \gamma_{kl} k_{jt} l_{jt} + \gamma_{km} k_{jt} m_{jt} + \gamma_{lm} l_{jt} m_{jt} \end{aligned}$$

to estimate the intermediate input elasticity and

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

for the constant of integration.

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 datasets.

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

[39] See also ACF for a discussion of serially uncorrelated shocks in the context of a value-added production function.

[40] 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 et al., 2016).

7.1 Monte Carlo Evidence on Estimator Performance

Under Assumptions 1-5, our procedure generates nonparametric estimates of the production function. In order 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 a Cobb-Douglas production function, using the same basic setup as described in Section 3.1.1 and the Appendix. (See Online Appendix O4 for additional details.)

The first two columns of Table 1 summarize the results of estimating the production function using our nonparametric procedure on 100 simulated datasets. The data is generated under a constant output elasticity of intermediate inputs and capital of 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.

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 elasticties implied by the Cobb-Douglas specification. Finally, none of the estimated elasticities are either below 0 or above 1.

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] 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

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

investigate the robustness of our estimator to violations of this assumption. In order to do so, in Online Appendix O4, we discuss 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 datasets, on average, pay substantial adjustment costs for intermediate inputs of almost 10% of the value of total gross output.

We generate 100 Monte Carlo samples for each of 9 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 to 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 in the Online Appendix. 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 though, our procedure continues to perform quite well, even for the largest values of adjustment costs, with our estimates reflecting only a small bias in the elasticities.

7.2 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 dataset we use comes from the Colombian manufacturing census covering all manufacturing plants with more than 10 employees from 1981-

1991. This dataset has been used in several studies, including Roberts and Tybout (1997), Clerides et al. (1998), and Das et al. (2007). The second dataset comes from the census of Chilean manufacturing plants conducted by Chile's Instituto Nacional de Estadística (INE). It covers all firms from 1979-1996 with more than 10 employees. This dataset has also been used extensively in previous studies, both in the production function estimation literature (LP) and in the international trade literature (Pavcnik, 2002 and Alvarez and López, 2005).[42]

We estimate separate production functions for the five largest 3-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 2 for both the elasticity and the integration constant in the production function.[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 estimates both 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 that are precisely estimated, as evidenced by the low standard errors. Intermediate inputs have the highest elasticity, with an average ranging from 0.50-0.67, 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, are also sensible, ranging from 0.99-1.15. Food Products (311) and Textiles (321) are the most capital intensive industries in Colombia, and in Chile the most capital intensive 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 to the aggregate specification denoted "All" in

[42] We construct the variables adopting the convention used by Greenstreet (2007) with the Chilean dataset, and employ the same approach with the Colombian dataset. 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 collar 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 et al. (2003).

[43] For all of the estimates 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.

the tables.

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 the industries with the smallest number of observations.

In order to evaluate the importance of transmission bias, we compare estimates from our procedure to 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 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 OP, 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 of each panel 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 to 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 to OLS, and 16% for Chile. The OLS estimates also imply smaller levels of persistence in productivity over time. The

[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.

average correlation coefficient between current and lagged productivity is 0.64 for our estimates and 0.53 under OLS.

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 non-exporters 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 based on advertising expenditures, not only are there sizeable differences in average productivities between OLS and our estimates, but in many cases the sign of the relationship actually changes. When compare 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% for Chile.

7.2.1 Robustness Checks and Extensions

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 been commonly assumed to be flexible in the literature. We believe that this is a reasonable assumption because a) the model period is typically a 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. 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 non-flexible. In particular, the production function we estimate is of the form $F\left(k_{jt}, l_{jt}, rm_{jt}, ns_{jt}\right) e^{\omega_{jt}+\varepsilon_{jt}}$, where rm denotes raw materials and ns denotes energy plus services. In one specification we assume rm to be non-flexible

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

and ns to be flexible, and in the other specification we assume the opposite. See Online Appendix O6 for these results. Overall the results are sensible and the comparison to OLS is similar to our main results.

Fixed Effects As we detail in Online Appendix O5, our identification and estimation strategy can be easily extended to incorporate fixed effects in the production function.[47] The production function allowing for fixed effects, a_j, can be written as $Y_{jt} = F(k_{jt}, l_{jt}, m_{jt}) e^{a_j + \omega_{jt} + \varepsilon_{jt}}$.[48] A common drawback of models with 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 GM).

In Online Appendix O6, 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.

Extra Unobservables As we show in Online Appendix O5, 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 Online Appendix O6 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.

Relaxing Independence of the Ex-Post Shock In order to investigate the importance of our assumption that $\mathcal{E} = E[e^{\varepsilon_{jt}}]$ is a constant, in Online Appendix O6 we present estimates in which we

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

[48] See Kasahara et al. (2015) for an important extension of our approach to the general case of firm-specific production functions.

allow $E\left[e^{\varepsilon_{jt}} \mid \mathcal{I}_{jt}\right]$ 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-squared values around 1%. As shown in Tables O6.9 and O6.10, the results are overall very similar to our baseline estimates in Tables 2 and 3.

8 Conclusion

In this paper we show new results regarding the nonparametric identification of gross output production functions in the presence of both flexible and non-flexible 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 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 datasets, we show that our nonparametric estimator provides reasonable estimates of the production function elasticities. When we compare our estimates to 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, "Econometric Tools For Analyzing Market Outcomes," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Amsterdam: Elsevier, 2007, pp. 4171–4276.

_ , Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Alvarez, Roberto and Ricardo A. López, "Exporting and Performance: Evidence from Chilean Plants," *Canadian Journal of Economics*, 2005, *38* (4), 1384–1400.

Amiti, Mary and Jozef Konings, "Trade Liberalization, Intermediate Inputs, and Productivity: Evidence from Indonesia," *The American Economic Review*, 2007, *97* (5), 1611–1638.

Arellano, Manuel and Stephen Bond, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations," *Review of Economic Studies*, April 1991, *58* (2), 277–297.

Baily, Martin N., Charles Hulten, and David Campbell, "Productivity Dynamics in Manufacturing Plants," *Brookings Papers on Economic Activity. Microeconomics*, 1992, pp. 187–267.

Bartelsman, Eric J. and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," Finance and Economics Discussion Series 2000-2019, Board of Governors of the Federal Reserve System (U.S.) 2000.

_ and _ , "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, April 1997, *105* (2), 249 – 283.

Bergoeing, Raphael, Andrés Hernando, and Andrea Repetto, "Idiosyncratic Productivity Shocks and Plant-Level Heterogeneity," Documentos de Trabajo 173, Centro de Economía Aplicada, Universidad de Chile 2003.

Bernard, Andrew B. and J. Bradford Jensen, "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-1987," *Brookings Papers on Economic Activity. Microeconomics*, 1995, pp. 67–119.

_ and _ , "Exceptional Exporter Performance: Cause, Effect or Both?," *Journal of International Economics*, 1999, *47* (1), 1–25.

_ , Jonathan Eaton, J. Bradford Jensen, and Samuel Kortum, "Plants and Productivity in International Trade," *American Economic Review*, 2003, *93* (4), 1268–1290.

Bils, Mark, Pete Klenow, and Cian Ruane, "Misallocation or Mismeasurement?," 2017. Working Paper.

Blum, Bernardo S., Sebastian Claro, Ignatius Horstmann, and David A. Rivers, "The Effects of Foreign and Domestic Demand Heterogeneity on Firm Dynamics with Implications to Aggregate Productivity and Trade," 2017. Working Paper.

Blundell, Richard and Stephen Bond, "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models," *Journal of Econometrics*, August 1998, *87* (1), 115–143.

_ **and** _ , "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews*, 2000, *19* (3), 321–340.

Bond, Stephen and Måns Söderbom, "Adjustment Costs and the Identification of Cobb Douglas Production Functions," 2005. Unpublished Manuscript, The Institute for Fiscal Studies, Working Paper Series No. 05/4.

Bruno, Michael, "Duality, Intermediate Inputs, and Value-Added," in M. Fuss and McFadden D., eds., *Production Economics: A Dual Approach to Theory and Applications*, Vol. 2, Amsterdam: North-Holland, 1978, chapter 1.

Caves, Douglas W., Laurits R. Christensen, and W. Erwin Diewert, "The Economic Theory of Index Numbers and the Measurement of Input, Output, and Productivity," *Econometrica*, November 1982, *50* (6), 1393–1414.

Chen, Xiaohong, "Large Sample Sieve Estimation of Semi-Nonparametric Models," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Amsterdam: Elsevier, 2007, pp. 5549–5632.

Clerides, Sofronis K., Saul Lach, and James R. Tybout, "Is Learning by Exporting Important? Micro-Dynamic Evidence from Colombia, Mexico, and Morocco," *The Quarterly Journal of Economics*, 1998, *113* (3), 903–947.

Collard-Wexler, Allan, "Productivity Dispersion and Plant Selection in the Ready-Mix Concrete Industry," 2010. NYU Stern working paper.

Das, Sanghamitra, Mark J. Roberts, and James R. Tybout, "Market Entry Costs, Producer Heterogeneity, and Export Dynamics," *Econometrica*, May 2007, *75* (3), 837–873.

Dhrymes, Phoebus J., *The Structure of Production Technology Productivity and Aggregation Effects*, US Department of Commerce, Bureau of the Census, 1991.

Doraszelski, Ulrich and Jordi Jaumandreu, "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 2013, *80* (4), 1338–1383.

_ **and** _ , "Measuring the Bias of Technological Change," 2015. Working Paper.

Foster, Lucia, John Haltiwanger, and Chad Syverson, "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?," *American Economic Review*, March 2008, *98* (1), 394–425.

Fox, Jeremy and Valérie Smeets, "Does Input Quality Drive Measured Differences in Firm Productivity?," *International Economic Review*, November 2011, *52* (4), 961–989.

Gandhi, Amit, Salvador Navarro, and David A. Rivers, "How Heterogeneous is Productivity? A Comparison of Gross Output and Value Added," 2017. Working Paper.

Goldberger, Arthur S., "The Interpretation and Estimation of Cobb-Douglas Functions," *Econometrica*, July-October 1968, *35* (3-4), 464–472.

Greenstreet, David, "Exploiting Sequential Learning to Estimate Establishment-Level Productivity Dynamics and Decision Rules," Economics Series Working Papers 345, University of Oxford, Department of Economics 2007.

Grieco, Paul, Shengyu Li, and Hongsong Zhang, "Production Function Estimation with Unobserved Input Price Dispersion," *International Economic Review*, 2016, *57* (2), 665–690.

Griliches, Zvi and Jacques Mairesse, "Production Functions: The Search for Identification," in "Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium," New York: Cambridge University Press, 1998, pp. 169–203.

_ **and Vidar Ringstad**, *Economies of Scale and the Form of the Production Function: An Econometric Study of Norwegian Manufacturing Establishment Data*, North-Holland Pub. Co. (Amsterdam), 1971.

Hahn, Jinyong, Zhipeng Liao, and Geert Ridder, "Nonparametric Two-Step Sieve M Estimation and Inference," 2016. Working Paper.

Halpern, László, Miklós Koren, and Adam Szeidl, "Imported Inputs and Productivity," *The American Economic Review*, 2015, *105* (12), 3660–3703.

_ **and Kenneth J. Singleton**, "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, September 1982, *50* (5), 1269–1286.

_ **and Thomas J. Sargent**, "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control*, 1980, *2*, 7–46.

Heckman, James J., "Shadow Prices, Market Wages, and Labor Supply," *Econometrica*, July 1974, *42* (4), 679–694.

_ **and Edward J. Vytlacil**, "Econometric Evaluation of Social Programs, Part II: Using the Marginal Treatment Effect to Organize Alternative Econometric Estimators to Evaluate Social Programs, and to Forecast Their Effects in New Environments," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Elsevier, 2007, pp. 4875–5143.

_ **and Richard Robb**, "Alternative Methods for Evaluating the Impact of Interventions," in James J. Heckman and Burton Singer, eds., *Longitudinal Analysis of Labor Market Data*, Cambridge University Press, New York, 1985.

_ **, Rosa L. Matzkin, and Lars Nesheim**, "Nonparametric Identification and Estimation of Nonadditive Hedonic Models," *Econometrica*, 2010, *78* (5), 1569–1591.

Horowitz, Joel L, "The Bootstrap," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 5, Amsterdam: Elsevier, 2001, pp. 3159–3228.

Houthakker, Hendrik S., "Revealed Preference and the Utility Function," *Economica*, May 1950, *17* (66), 159–174.

Judd, Kenneth L, *Numerical Methods in Economics*, Cambridge: MIT Press, 1998.

Kasahara, Hiroyuki and Joel Rodrigue, "Does the Use of Imported Intermediates Increase Productivity? Plant-level Evidence," *The Journal of Development Economics*, 2008, *87* (1), 106–118.

_ , **Paul Schrimpf, and Michio Suzuki**, "Identification and Estimation of Production Function with Unobserved Heterogeneity," 2015. Working Paper.

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

Klein, Roger and Francis Vella, "Estimating a Class of Triangular Simultaneous Equations Models Without Exclusion Restrictions," *Journal of Econometrics*, 2010, *154* (2), 154–164.

Koopmans, Tjalling C., Herman Rubin, and Roy Bergh Leipnik, "Measuring the Equation Systems of Dynamic Economics," in "Statistical Inference in Dynamic Economic Models," Vol. 10, Wiley New York, 1950.

Laffont, Jean-Jacques and Quang Vuong, "Structural Analysis of Auction Data," *American Economic Review, P&P*, 1996, *86* (2), 414–420.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, April 2003, *70* (2), 317–342.

Lewbel, Arthur, "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models," *Journal of Business & Economic Statistics*, 2012, *30* (1), 67–80.

_ , "Detecting Learning by Exporting," *American Economic Journal: Microeconomics*, 2013, *5* (3), 1–21.

_ **, Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik**, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Lucas, Robert E. and Thomas J. Sargent, *Rational Expectations and Econometric Practice*, University of Minnesota Press, 1981.

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

Nerlove, Marc, "Returns to Scale in Electricity Supply," in Christ et. al., ed., *Measurement in Economics*, Stanford: Stanford University Press, 1963, pp. 167–200.

_ **and James L. Powell**, "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 2003, *71* (5), 1565–1578.

_ **,** _ **, and F Vella**, "Nonparametric Estimation of Triangular Simultaneous Equations Models," *Econometrica*, 1999, *67* (3), 565–603.

Oberfield, Ezra, "Productivity and Misallocation During a Crisis: Evidence from the Chilean Crisis of 1982," *Review of Economic Dynamics*, 2013, *16* (1), 100–119.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996, *64* (6), 1263–1297.

Paarsch, Harry J., "Deciding Between the Common and Private Value Paradigms in Empirical Models of Auctions," *Journal of Econometrics*, 1992, *51* (1-2), 191–215.

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

Pombo, Carlos, "Productividad Industrial en Colombia: Una Aplicacion de Numeros Indicey," *Revista de Economia del Rosario*, 1999, *2* (3), 107–139.

Pozzi, Andrea and Fabiano Schivardi, "Demand or Productivity: What Determines Firm Growth?," *The RAND Journal of Economics*, 2016, *47* (3), 608–630.

Rigobon, Roberto, "Identification Through Heteroskedasticity," *Review of Economics and Statistics*, 2003, *85* (4), 777–792.

Roberts, Mark J. and James R. Tybout, "The Decision to Export in Colombia: An Empirical Model of Entry with Sunk Costs," *American Economic Review*, 1997, *87* (4), 545–564.

Sargent, Thomas J., "Rational Expectations, Econometric Exogeneity, and Consumption," *Journal of Political Economy*, 1978, *86* (4), 673–700.

Solow, Robert M., "Technical Change and the Aggregate Production Function," *Review of Economics and Statistics*, 1957, *39* (3), 312–320.

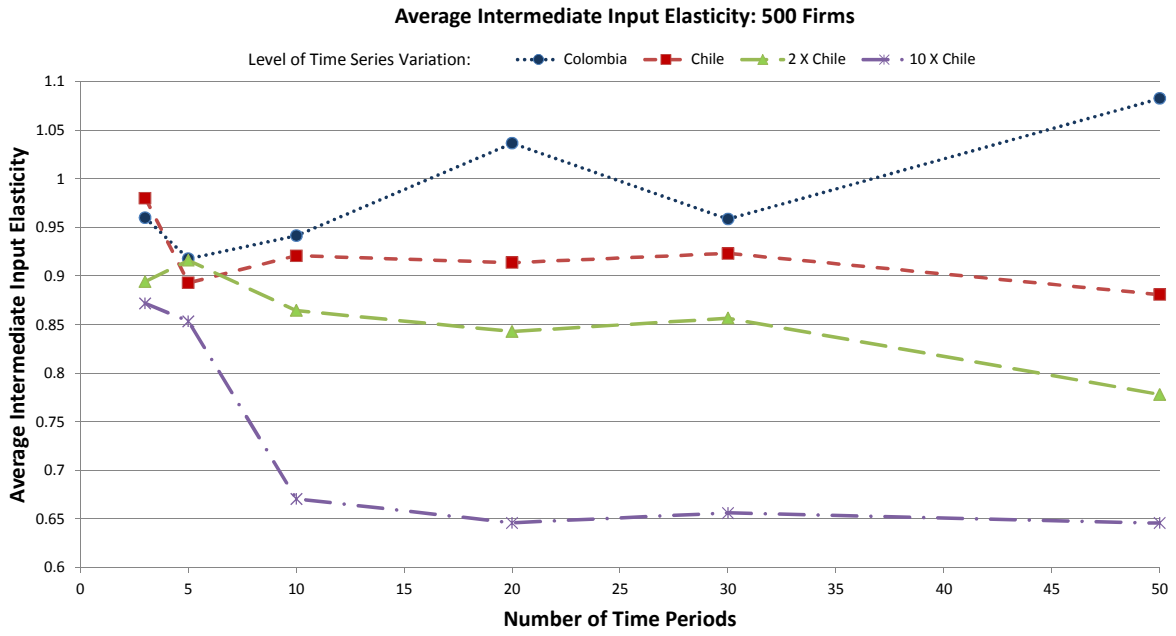
Stone, Richard, "Linear Expenditure Systems and Demand Analysis: An Application to the Pattern of British Demand," *The Economic Journal*, September 1954, *64* (255), 511–527.

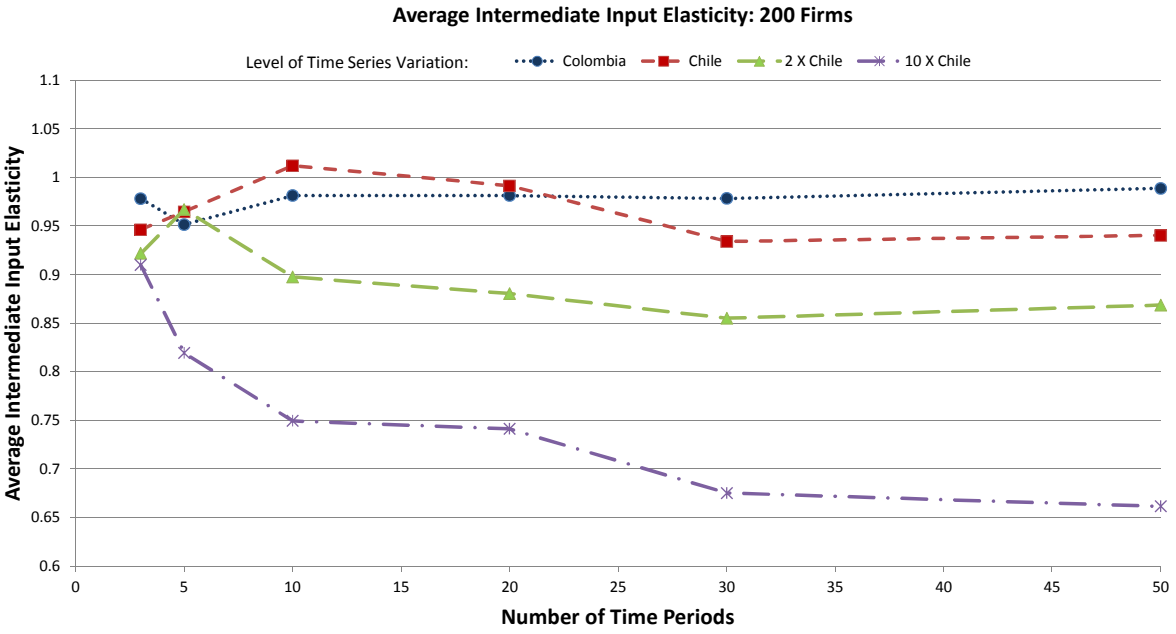
Syverson, Chad, "Product Substitutability and Productivity Dispersion," *The Review of Economics and Statistics*, 2004, *86* (2), 534–550.

_ , "What Determines Productivity?," *Journal of Economic Literature*, June 2011, *49* (2), 326–365.

Wooldridge, Jeffrey M., "On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables," *Economics Letters*, September 2009, *104* (3), 112–114.

Figure 1: Monte Carlo--Proxy Variable Estimator Applied to Gross Output





Notes: This figure presents the results from applying a proxy variable estimator extended to gross output to Monte Carlo data generated as described in Online 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 average of the estimated intermediate input elasticity across 100 Monte Carlo simulations. The true value of the elasticity is 0.65. The first panel includes 500 firms, and the second inlcudes 200.

Table 1: Monte Carlo--GNR Estimator Performance

	True Functional Form					
	Cobb Douglas		**CES**		**Translog**	
	At True Parameters	GNR Estimates	At True Parameters	GNR Estimates	At True Parameters	GNR Estimates
Intermediates						
Average Mean Elasticity	0.6500	0.6502	0.6747	0.6746	0.6574	0.6572
	--	(0.0015)	(0.0027)	(0.0030)	(0.0007)	(0.0014)
Average St. Dev.	0	0.0038	0.1197	0.1193	0.0321	0.0324
	--	(0.0012)	(0.0014)	(0.0018)	(0.0004)	(0.0012)
Average Fraction Outside of (0,1)	0	0.0000	0	0.0000	0.0000	0.0000
	--	(0.0000)	--	(0.0000)	(0.0000)	(0.0000)
Capital						
Average Mean Elasticity	0.2500	0.2492	0.2253	0.2196	0.2263	0.2251
	--	(0.0063)	(0.0027)	(0.0066)	(0.0006)	(0.0075)
Average St. Dev.	0	0.0086	0.1197	0.1209	0.0333	0.0347
	--	(0.0041)	(0.0014)	(0.0022)	(0.0004)	(0.0018)
Average Fraction Outside of (0,1)	0	0.0000	0	0.0090	0.0000	0.0000
	--	(0.0000)	--	(0.0047)	(0.0000)	(0.0000)
Sum						
Average Mean Elasticity	0.9000	0.8994	0.9000	0.8942	0.8836	0.8823
	--	(0.0064)	--	(0.0067)	(0.0002)	(0.0074)
Average St. Dev.	0	0.0090	0	0.0221	0.0056	0.0114
	--	(0.0044)	--	(0.0034)	(0.0001)	(0.0061)

Notes:

a. In this table we compare estimates of the production function elasticities using our nonparametric procedure (GNR) to the true values. We simulate data from thee different parametric production functions: Cobb-Douglas, CES, and translog. See Online Appendix O4 for the details.

b. 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 2 for each of the two nonparametric functions (D and ℓ) of our approach.

c. For each simulated dataset, 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). In the table we report the average of these three statistics across each of the simulated datasets, as well as the corresponding standard error (calculated across the 100 simulations).

d. 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.

e. For cases in which there is no variation in a statistic across simulations under the true parameter values, we report "--" for the associated standard error. For example, under Cobb-Douglas, the true production function elasticites 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 "0" with no decimals. For example, under CES, the elasticities are always strictly positive and less than 1 given our chosen parameter values.

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	0.22	0.15	0.32	0.21	0.42	0.32	0.44	0.32	0.43	0.29	0.35	0.26
	(0.02)	(0.01)	(0.03)	(0.02)	(0.02)	(0.01)	(0.05)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Capital	0.12	0.04	0.16	0.06	0.05	0.01	0.04	0.03	0.10	0.03	0.14	0.06
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Intermediates	0.67	0.82	0.54	0.76	0.52	0.68	0.51	0.65	0.53	0.73	0.54	0.72
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
Sum	1.01	1.01	1.01	1.03	0.99	1.01	0.99	1.00	1.06	1.05	1.04	1.04
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.04)	(0.02)	(0.01)	(0.01)	(0.00)	(0.00)
Mean(Capital) / Mean(Labor)	0.55	0.27	0.49	0.27	0.12	0.04	0.08	0.08	0.23	0.11	0.40	0.23
	(0.08)	(0.07)	(0.09)	(0.06)	(0.04)	(0.02)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)	(0.01)
Chile												
Labor	0.28	0.17	0.45	0.26	0.45	0.29	0.40	0.20	0.52	0.32	0.38	0.20
	(0.01)	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)
Capital	0.11	0.05	0.11	0.06	0.06	0.03	0.07	0.02	0.13	0.07	0.16	0.09
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Intermediates	0.67	0.83	0.54	0.75	0.56	0.74	0.59	0.81	0.50	0.71	0.55	0.77
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
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
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
Mean(Capital) / Mean(Labor)	0.39	0.28	0.24	0.22	0.14	0.12	0.18	0.12	0.25	0.21	0.43	0.42
	(0.03)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.05)	(0.03)	(0.04)	(0.02)	(0.02)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers in the first column are based on a gross output specification using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ) of our approach (labeled 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 2 with OLS.

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, and intermediate input elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

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
	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.04)	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)
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
	(0.05)	(0.02)	(0.07)	(0.04)	(0.03)	(0.02)	(0.12)	(0.06)	(0.03)	(0.02)	(0.02)	(0.01)
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
	(0.08)	(0.05)	(0.14)	(0.08)	(0.05)	(0.04)	(0.22)	(0.15)	(0.06)	(0.04)	(0.03)	(0.02)
Exporter	0.14	0.09	0.02	-0.01	0.05	0.00	0.15	0.10	0.08	0.03	0.06	0.01
	(0.05)	(0.04)	(0.03)	(0.01)	(0.03)	(0.01)	(0.14)	(0.09)	(0.03)	(0.02)	(0.01)	(0.01)
Importer	0.04	-0.02	0.05	0.00	0.12	0.02	0.05	-0.03	0.10	0.05	0.11	0.04
	(0.02)	(0.01)	(0.04)	(0.01)	(0.03)	(0.01)	(0.08)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Advertiser	-0.03	-0.07	0.08	-0.04	0.05	-0.03	0.04	-0.02	0.05	0.00	0.03	-0.02
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Wages > Median	0.09	0.06	0.18	0.10	0.18	0.13	0.15	0.11	0.22	0.13	0.20	0.13
	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.01)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
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
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
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
	(0.02)	(0.01)	(0.05)	(0.04)	(0.05)	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)	(0.02)	(0.01)
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
	(0.05)	(0.02)	(0.09)	(0.07)	(0.09)	(0.05)	(0.11)	(0.07)	(0.10)	(0.06)	(0.04)	(0.03)
Exporter	0.02	-0.01	0.02	-0.02	0.09	0.01	0.00	-0.02	-0.01	0.00	0.03	-0.01
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)
Importer	0.14	0.03	0.10	0.04	0.14	0.06	0.15	0.07	0.11	0.06	0.15	0.09
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)	(0.03)	(0.02)	(0.02)	(0.01)	(0.01)
Advertiser	0.04	0.00	0.04	0.01	0.06	0.02	0.03	0.01	0.01	0.01	0.06	0.04
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Wages > Median	0.21	0.12	0.19	0.15	0.22	0.16	0.21	0.13	0.22	0.16	0.30	0.24
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers in the first column are based on a gross output specification using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and $\mathcal{E}$) of our approach (labeled 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 2 with OLS.

c. 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) based on 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.

Appendix: Monte Carlo Setup

In this appendix we describe the general structure of our simulated data, which we then vary depending on the Monte Carlo experiment. We consider a panel of (up to) 500 firms over (up to) 50 periods, and repeat the simulations (up to) 500 times. To simplify the problem we abstract away from labor and consider the following Cobb-Douglas production function

$$Y_{jt} = K_{jt}^{\alpha_k} M_{jt}^{\alpha_m} e^{\omega_{jt}+\varepsilon_{jt}},$$

where $\alpha_k = 0.25$, $\alpha_m = 0.65$, and ε_{jt} is measurement error that is distributed $N(0, 0.07)$. ω_{jt} follows an AR(1) process

$$\omega_{jt} = \delta_0 + \delta\omega_{jt-1} + \eta_{jt},$$

where $\delta_0 = 0.2$, $\delta = 0.8$, and $\eta_{jt} \sim N(0, 0.04)$. We select the variances of the errors and the AR(1) parameters to roughly correspond to the estimates from our Chilean and Colombian datasets.

The environment facing the firms is the following. At the beginning of each period, firms choose investment I_{jt} and intermediate inputs M_{jt}. Investment determines the next period's capital stock via the law of motion for capital

$$K_{jt+1} = (1 - \kappa_j) K_{jt} + I_{jt},$$

where $\kappa_j \in \{0.05, 0.075, 0.10, 0.125, 0.15\}$ is the depreciation rate which is distributed uniformly across firms. Depending on the simulation, intermediate inputs may be subject to quadratic adjustment costs of the form

$$C_{jt}^M = 0.5b\frac{(M_{jt} - M_{jt-1})^2}{M_{jt}},$$

where $b \in [0, 1]$ is a parameter that indexes the level of adjustment costs, which we vary in our simulations. The case of $b = 0$ corresponds to intermediate inputs m_{jt} being chosen flexibly in period t.

Firms choose investment and intermediate inputs to maximize expected discounted profits. The

problem of the firm, written in recursive form, is thus given by

$$\begin{aligned} V\left(K_{jt}, M_{jt-1}, \omega_{jt}\right) &= \max_{I_{jt}, M_{jt}} P_t K_{jt}^{\alpha_k} M_{jt}^{\alpha_m} e^{\omega_{jt}} - P_t^I I_{jt} - \rho_t M_{jt} \\ &\quad -0.5b\frac{\left(M_{jt} - M_{jt-1}\right)^2}{M_{jt}} + \beta E_t V\left(K_{jt+1}, M_{jt}, \omega_{jt+1}\right) \\ s.t. & \\ & K_{jt+1} = \left(1 - \kappa_j\right) K_{jt} + I_{jt} \\ & I_{jt} \geq 0, M_{jt} \geq 0 \\ & \omega_{jt+1} = \delta_0 + \delta\omega_{jt} + \eta_{jt+1}. \end{aligned}$$

The price of output P_t and the price of intermediate inputs ρ_t are set to 1. The price of investment P_t^I is set to 8, and there are no other costs to investment. The discount factor is set to 0.985.

In order for our Monte Carlo simulations not to depend on the initial distributions of (k, m, ω), we simulate each firm for a total of 200 periods, dropping the first 150 periods. The initial conditions, k_1, m_0, and ω_1 are drawn from the following distributions: $U\left(11, 400\right)$, $U\left(11, 400\right)$, and $U\left(1, 3\right)$. Since the firm's problem does not have an analytical solution, we solve the problem numerically by value function iteration with an intermediate modified policy iteration with 100 steps, using a multi-linear interpolant for both the value and policy functions.[49]

[49] See Judd (1998) for details.

Online Appendix O1: Extensions of Theorem 1

In what follows we show that the results of Theorem 1 can be extended to the cases in which 1) dynamic panel data methods are used and 2) investment is used as the proxy instead of intermediate inputs.

Dynamic Panel Methods Equation (25) in Section 6.3 sets up dynamic panel methods under the common AR(1) assumption on ω in terms of the following conditional moment restriction:

$$E\ \ y_{jt} \mid \Gamma_{jt}^{DP}\ \ = E\ \ f\left(k_{jt}, l_{jt}, m_{jt}\right) + \delta_0 + \delta\left(y_{jt-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \mid \Gamma_{jt}^{DP}\ \ , \tag{26}$$

where $\Gamma_{jt}^{DP} = \Gamma_{jt} \setminus y_{jt-1}$.[50] One difference from the proxy variable method is that there is no first stage, and everything is based on the analogue of the second stage, i.e., the functional restriction in equation (26).

Theorem 4. *In the absence of time series variation in relative prices, $d_t = d \forall t$, under the model defined by Assumptions 1 - 4 and assuming an AR(1) process for ω, there exists a continuum of alternative $\tilde{f}, \tilde{h}$ defined by*

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

for any $a \in (0,1)$ that satisfy the same functional restriction (26) *as the true (f^0, h^0).*

Proof. We begin by noting that under the AR(1), $h^0\left(x\right) = \delta_0^0 + \delta^0 x$ and $\tilde{h}\left(x\right) = \tilde{\delta}_0 + \tilde{\delta} x$, where $\tilde{\delta}_0 = ad\left(1-\delta^0\right) + \left(1-a\right)\delta_0^0$ and $\tilde{\delta} = \delta$. Next, given the definition of $\tilde{f}, \tilde{h}$ and noting that

[50] Note that y_{jt-1} needs to be excluded from the conditioning set since by definition it is correlated with ε_{jt-1}, which is part of ψ_{jt}, the error term in y_{jt}.

$d_t = d\,\forall t$, we have

$$\begin{aligned}
\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) + \tilde{\delta}_0 + \tilde{\delta}\ \ y_{jt-1} - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \quad &= \\
f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + a\ \ \mathbb{M}^0\ ^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ \ ad\ \ 1 - \delta^0\ \ + (1-a)\,\delta_0^0 & \\
+\delta^0\ \ y_{jt-1} - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) - \ \ \mathbb{M}^0\ ^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \quad &= \\
f^0\left(k_{jt}, l_{jt}, m_{jt}\right) + \delta_0^0 + \delta^0\ \ y_{jt-1} - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) & \\
+a \quad \mathbb{M}^0\ ^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d - \delta_0^0 - \delta^0 \quad \mathbb{M}^0\ ^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d \quad . &
\end{aligned}$$

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

$$\begin{aligned}
E\ \ \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{DP}\ \ + \tilde{\delta}_0 + \tilde{\delta}\ \ E\ \ y_{jt-1} \mid \Gamma_{jt}^{DP}\ \ - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \quad &= \\
E\ \ f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{DP}\ \ + \delta_0^0 + \delta^0\ \ E\ \ y_{jt-1} \mid \Gamma_{jt}^{DP}\ \ - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) & \\
+a\ \ E \quad \mathbb{M}^0\ ^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d \mid \Gamma_{jt}^{DP}\ \ - \delta_0^0 - \delta^0 \quad \mathbb{M}^0\ ^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d \quad &= \\
E\ \ f^0\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{DP}\ \ + \delta_0^0 + \delta^0\ \ E\ \ y_{jt-1} \mid \Gamma_{jt}^{DP}\ \ - f^0\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \ \ . &
\end{aligned}$$

The last equality follows from the observation that $\left(\mathbb{M}^0\right)^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d = \omega_{jt}$ and $E\ \ \omega_{jt} \mid \Gamma_{jt}^{DP}\ \ = \delta_0^0 + \delta^0 \omega_{jt-1}$. Thus $\left(f^0, h^0\right)$ and $\ \ \tilde{f}, \tilde{h}\ \ $ satisfy the functional restriction and cannot be distinguished via instrumental variables. □

Investment as the Proxy Using investment as the proxy variable requires an analogue of Assumption (3) for investment.

Assumption 6. *Investment in physical capital, denoted i_{jt}, is assumed strictly monotone in a single unobservable ω_{jt}:*

$$i_{jt} \ \ = \ \ \mathbb{I}_t\left(k_{jt}, l_{jt}, \omega_{jt}\right). \tag{27}$$

Using investment, the first stage of the proxy variable procedure applied to gross output would recover

$$E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, i_{jt}\right] = f\left(k_{jt}, l_{jt}, m_{jt}\right) + E\left[\omega_{jt} \mid k_{jt}, l_{jt}, m_{jt}, i_{jt}\right]. \tag{28}$$

Under Assumption 6, $\omega_{jt} = \mathbb{I}_t^{-1}\left(k_{jt}, l_{jt}, i_{jt}\right)$ and under Assumption 3, $\omega_{jt} = \mathbb{M}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t$, and therefore $i_{jt} = \tilde{\mathbb{I}}_t\left(k_{jt}, l_{jt}, m_{jt}\right)$. This implies that we can rewrite the first stage in equation (28)

as:

$$E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, i_{jt}\right] = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathbb{I}_t^{-1}\left(k_{jt}, l_{jt}, i_{jt}\right) \equiv \phi_t^i\left(k_{jt}, l_{jt}, m_{jt}, i_{jt}\right).$$

But we can also write it as

$$E\left[y_{jt} \mid k_{jt}, l_{jt}, m_{jt}, i_{jt}\right] = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathbb{M}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t = \phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t,$$

where notice that i_{jt} has dropped out, and the first stage corresponds exactly to the case of using intermediate inputs as the proxy. Therefore we have that

$$\phi_t^i\left(k_{jt}, l_{jt}, m_{jt}, i_{jt}\right) = \phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d_t. \tag{29}$$

This leads to an analogue of the functional restriction (9) given by

$$\begin{aligned} E\ \ y_{jt} \mid \Gamma_{jt}^i\ &= E\ \ f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^i\ \ + h\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1} - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right) \\ &= E\ \ f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^i\ \ + h\ \ \phi_t^i\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\ \ . \end{aligned} \tag{30}$$

Theorem 5. *In the absence of time series variation in relative prices, $d_t = d\ \forall t$, under the model defined by Assumptions 1 - 4 and 6, there exists a continuum of alternative $\tilde{f}, \tilde{h}$ defined by*

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

for any $a \in (0,1)$ that satisfy the same functional restriction (30) *as the true (f^0, h^0).*

Proof. Let $\Gamma_{jt}^{i} = \Gamma_{jt} \cup \{i_{jt}, ...i_{j1}\}$. Given the definition of $\tilde{f}, \tilde{h}$, we have

$$\tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) + \tilde{h} \quad \phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - \tilde{f}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \quad =$$

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

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

$$\begin{array}{c} f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right) + a\left(\phi\left(k_{jt}, l_{jt}, m_{jt}\right) + d - f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right)\right) \\ + (1-a)\, h^{0}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d - f^{0}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{array}$$

Given equation (29), this can be re-written as

$$\begin{array}{c} f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right) + a\left(\phi_{t}^{i}\left(k_{jt}, l_{jt}, m_{jt}, i_{jt}\right) - f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right)\right) \\ + (1-a)\, h^{0}\left(\phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f^{0}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{array} \tag{31}$$

Next notice that for any (f, h) that solve the functional restriction (30), it must be the case that

$$E\ \ y_{jt} \mid \Gamma_{jt}^{i}\ \ = E\ \ f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{i}\ \ + h\ \ \phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\ \ .$$

Furthermore, from the definition of ϕ^{i}, it also follows that

$$E\ \ y_{jt} \mid \Gamma_{jt}^{i}\ \ = E\ \ \phi_{t}^{i}\left(k_{jt}, l_{jt}, m_{jt}, i_{jt}\right) \mid \Gamma_{jt}^{i}\ \ .$$

Hence,

$$\begin{array}{c} E\ \ \phi_{t}^{i}\left(k_{jt}, l_{jt}, m_{jt}, i_{jt}\right) - f\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{i}\ \ = \\ h\left(\phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\right). \end{array} \tag{32}$$

Now, take the conditional expectation of equation (31) (with respect to Γ_{jt}^{i})

$$\begin{array}{c} E\ \ f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{i}\ \ + ah^{0}\ \ \phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f^{0}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \\ + (1-a)\, h^{0}\ \ \phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f^{0}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) \quad = \\ E\ \ f^{0}\left(k_{jt}, l_{jt}, m_{jt}\right) \mid \Gamma_{jt}^{i}\ \ + h^{0}\ \ \phi_{t}^{i}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}, i_{jt-1}\right) - f^{0}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right)\ \ , \end{array}$$

where the first line applies the relation in equation (32) to equation (31). Thus (f^0, h^0) and $\tilde{f}, \tilde{h}$ satisfy the functional restriction and cannot be distinguished via instrumental variables. □

As discussed in Section 6.3, in the context of dynamic panel, it may be possible to relax the scalar unobservability / montonicity assumption on intermediate inputs. This is also the case for using investment in a proxy variable setup. A key difference for the case of investment as the proxy is that one must be careful that the way in which this assumption is relaxed does not also violate the scalar unobservability / monotonicity assumption for investment, Assumption (6).

Online Appendix O2: A Parametric Example

In order to further illustrate the mechanisms behind Theorem 1 and its corollaries, we consider a parametric example. Suppose that the true production function is Cobb-Douglas $F(k_{jt}, l_{jt}, m_{jt}) = K_{jt}^{\alpha_k} L_{jt}^{\alpha_l} M_{jt}^{\alpha_m}$, and productivity follows an AR(1) process $\omega_{jt} = \delta_0 + \delta\omega_{jt-1} + \eta_{jt}$. Replacing the first stage estimates of ϕ into the production function we obtain:

$$\begin{aligned} y_{jt} &= constant + \alpha_k k_{jt} + \alpha_l l_{jt} + \alpha_m m_{jt} \\ &\quad + \delta\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d_{t-1} - \alpha_k k_{jt-1} - \alpha_l l_{jt-1} - \alpha_m m_{jt-1}\right) + \eta_{jt} + \varepsilon_{jt}. \end{aligned}$$

If we plug in for m_{jt} using the first-order condition and combine constants we have

$$\begin{aligned} y_{jt} &= \tilde{constant} + \left(\frac{\alpha_k}{1-\alpha_m}\right) k_{jt} + \left(\frac{\alpha_l}{1-\alpha_m}\right) l_{jt} - \left(\frac{\alpha_m}{1-\alpha_m}\right) d_t + \left(\frac{\delta}{1-\alpha_m}\right) d_{t-1} \\ &\quad + \frac{\delta}{1-\alpha_m}\left(\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) - \alpha_k k_{jt-1} - \alpha_l l_{jt-1} - \alpha_m m_{jt-1}\right) + \left(\frac{1}{1-\alpha_m}\right)\eta_{jt} + \varepsilon_{jt}. \end{aligned}$$

Plugging in for the Cobb-Douglas parametric form of $\mathbb{M}^{-1}$, it can be shown that $\phi\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) = m_{jt-1} - \ln\alpha_m$, which implies

$$\begin{aligned} y_{jt} &= \tilde{constant} - \left(\frac{\alpha_m}{1-\alpha_m}\right) d_t + \left(\frac{\delta}{1-\alpha_m}\right) d_{t-1} + \left(\frac{\alpha_k}{1-\alpha_m}\right) k_{jt} + \left(\frac{\alpha_l}{1-\alpha_m}\right) l_{jt} \\ &\quad - \delta\left(\frac{\alpha_k}{1-\alpha_m}\right) k_{jt-1} - \delta\left(\frac{\alpha_l}{1-\alpha_m}\right) l_{jt-1} + \delta m_{jt-1} + \left(\frac{1}{1-\alpha_m}\right)\eta_{jt} + \varepsilon_{jt}. \end{aligned}$$

First consider the case in which there is no time series variation in d. From the equation above we can see that although variation in m_{jt-1} identifies δ, the coefficient on k_{jt} is equal to the coefficient on k_{jt-1} multiplied by $-\delta$, and the same is true for l. In other words, variation in k_{jt-1} and l_{jt-1} do not provide any additional information about the parameters of the production function. As a result, all we can identify is $\alpha_k \quad \frac{1}{1-\alpha_m}$ and $\alpha_l \quad \frac{1}{1-\alpha_m}$. To put it another way, the rank condition necessary for identification of this model is not satisfied.

In terms of our proposed alternative functions in Theorem 1, we would have

$$\tilde{\alpha}_k \;=\; (1-a)\,\alpha_k \;\;;\;\; \tilde{\alpha}_l \;=\; (1-a)\,\alpha_l \;\;;\;\; \tilde{\alpha}_m \;=\; (1-a)\,\alpha_m + a \;\;;\;\; \tilde{\delta} = \;\; \delta \;.$$

It immediately follows that $\frac{\tilde{\alpha}_k}{1-\tilde{\alpha}_m} \;=\; \frac{\alpha_k}{1-\alpha_m}$ and $\frac{\tilde{\alpha}_l}{1-\tilde{\alpha}_m} \;=\; \frac{\alpha_l}{1-\alpha_m}$, and thus our continuum of alternatives indexed by $a \in (0,1)$ satisfy the instrumental variables restriction.

For the case in which there is time series variation in d, this variation would identify α_m, and the model would be identified. However, as we discuss in the main body, relying on time series variation runs a risk of weak identification in applications. Doraszelski and Jaumandreu (2013) avoid this problem by exploiting observed cross-sectional variation in (lagged) prices as an instrument for identification. In contrast, our approach uses the first-order condition to form the share regression equation, which gives us a second structural equation that we use in identification and estimation. In terms of our Cobb-Douglas example, the second equation would be given by the following share equation $s_{jt} = \ln \mathcal{E} + \ln \alpha_m - \varepsilon_{jt}$. Given that $E[\varepsilon_{jt}] = 0$, $\{\ln \mathcal{E} + \ln \alpha_m\}$ is identified, therefore $\mathcal{E} = E[\exp(\{\ln \mathcal{E} + \ln \alpha_m\} - s_{jt})]$ is identified, and we can identify α_m.

Online Appendix O3: Extension of Theorem 1 Using Distributions

Following Hurwicz (1950) and using the language of Matzkin (2013), in what follows we define a structure as a distribution of the exogenous variables and a system of equations that generate the distribution of endogenous variables. In our case the endogenous variables are output and intermediate inputs, and these equations are the output and intermediate input demand equations. The functions f, h, and $\mathbb{M}$ are defined as features of the structure.

We now extend our result in Theorem 1 to show that absent time series variation in relative prices, $d_t = d \forall t$, the triple of unknown functions $\Theta = (f, h, \mathbb{M})$ cannot be identified from the full joint

distribution of the data

$$G_{y_{jT},m_{jT},k_{jT},l_{jT},...,y_{j2},m_{j2},k_{j2},l_{j2},y_{j1},m_{j1},k_{j1},l_{j1},d} = G_{y_{jT},m_{jT}|\Gamma_{jT}} \times ... \times G_{y_{j2},m_{j2}|\Gamma_{j2}} \times G_{\Gamma_{j2}},$$

where note that Γ_{j2} includes all period 1 variables.[51] The model described by Assumptions 1-4 imposes restrictions on $G_{y_{jt},m_{jt}|\Gamma_{jt}}$ for $t = 2, ..., T$. In what follows we show that one can generate an observationally equivalent structure that rationalizes $G_{y_{jt},m_{jt}|\Gamma_{jt}}$ for any arbitrary t, and therefore the triple $\Theta = (f, h, \mathbb{M})$ cannot be identified from the full joint distribution of the data.

For a given Θ, let

$$\varepsilon_{jt}^{\Theta} = y_{jt} - f(k_{jt}, l_{jt}, m_{jt}) - \mathbb{M}^{-1}(k_{jt}, l_{jt}, m_{jt}) - d,$$

and

$$\eta_{jt}^{\Theta} = \mathbb{M}^{-1}(k_{jt}, l_{jt}, m_{jt}) - h \ \ \mathbb{M}^{-1}(k_{jt-1}, l_{jt-1}, m_{jt-1}) + d \ \ .$$

In order to relate Θ to the (conditional) joint distribution of the data for an arbitrary period t, $G_{y_{jt},m_{jt}|\Gamma_{jt}}$, through the model, a joint distribution of the unobservables $G_{\eta_{jt}^{\Theta},\varepsilon_{jt}^{\Theta}|\Gamma_{jt}}$ needs to be specified. Let $E_G(\cdot)$ denote the expectation operator taken with respect to distribution G. We say that a triple $\Theta = (f, h, \mathbb{M})$ rationalizes the data if there exists a joint distribution $G_{\eta_{jt}^{\Theta},\varepsilon_{jt}^{\Theta}|\Gamma_{jt}} = G_{\eta_{jt}^{\Theta}|\Gamma_{jt}} \times G_{\varepsilon_{jt}^{\Theta}}$ that (i) generates the joint distribution $G_{y_{jt},m_{jt}|\Gamma_{jt}}$; (ii) satisfies the first stage moment restriction $E_{G_{\varepsilon_{jt}^{\Theta}}} \ \varepsilon_{jt}^{\Theta} \mid k_{jt}, l_{jt}, m_{jt} \ = 0$; (iii) satisfies the IV orthogonality restriction $E_{G_{\eta_{jt}^{\Theta},\varepsilon_{jt}^{\Theta}|\Gamma_{jt}}} \ \eta_{jt}^{\Theta} + \varepsilon_{jt}^{\Theta} \mid \Gamma_{jt} \ = 0$; and (iv) satisfies Assumption 3 (i.e., scalar unobservability and monotonicity of $\mathbb{M}$). Following Matzkin (2007), we say that, if there exists an alternative $\tilde{\Theta} \neq \Theta^0$ that rationalizes the data, then $\Theta^0 = (f^0, h^0, \mathbb{M}^0)$ is not identified from the joint distribution $G_{y_{jt},m_{jt}|\Gamma_{jt}}$ of the data.

Theorem 6. *Given the true $\Theta^0 = (f^0, h^0, \mathbb{M}^0)$, in the absence of time series variation in relative prices, $d_t = d \forall t$, under the model defined by Assumptions 1 - 4, there always exists a continuum of*

[51] As we also note in the main body, in our discussion before Theorem 1, one may not be interested in recovering h or $\mathbb{M}$. In our results below, regardless of whether h or $\mathbb{M}$ is identified, the production function f is not identified.

alternatives $\tilde{\Theta} \neq \Theta^0$*, defined by*

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

for any $a \in (0,1)$ *that exactly rationalize the data* $G_{y_{jt}, m_{jt}|\Gamma_{jt}}$.[52]

Proof. Let $\mathring{x}$ denote a particular value of the random variable x in its support. We first observe that, for any hypothetical $\Theta = (f, h, \mathbb{M})$, there always exists a distribution $G_{\eta_{jt}^{\Theta}, \varepsilon_{jt}^{\Theta}|\Gamma_{jt}}$ defined by

$$G_{\eta_{jt}^{\Theta}, \varepsilon_{jt}^{\Theta}|\Gamma_{jt}}\left(\mathring{\eta}_{jt}, \mathring{\varepsilon}_{jt} \mid \Gamma_{jt}\right) =$$

$$G_{y_{jt}, m_{jt}|\Gamma_{jt}}\left(\begin{array}{c} \left[\begin{array}{c} f\left(k_{jt}, l_{jt}, \mathbb{M}\left(k_{jt}, l_{jt}, h\left(\mathbb{M}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d\right) + \mathring{\eta}_{jt} - d\right)\right) \\ + h\left(\mathbb{M}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d\right) + \mathring{\eta}_{jt} + \mathring{\varepsilon}_{jt} \end{array}\right], \\ \mathbb{M}\left(k_{jt}, l_{jt}, h\left(\mathbb{M}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d\right) + \mathring{\eta}_{jt} - d\right) \end{array} \middle| \ \Gamma_{jt}\right)$$

that generates the conditional distribution of the data $G_{y_{jt}, m_{jt}|\Gamma_{jt}}$ through the model, hence (i) is satisfied.

Second, since the true model rationalizes the data, it follows that $E_{G_{\varepsilon_{jt}^{\Theta^0}}}\left[\ \varepsilon_{jt}^{\Theta^0} \mid k_{jt}, l_{jt}, m_{jt}\ \right] = 0$. The $\varepsilon_{jt}^{\tilde{\Theta}}$ implied by our alternative $\tilde{\Theta}$ is given by

$$\begin{aligned}
\varepsilon_{jt}^{\tilde{\Theta}} &= y_{jt} - \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) - \tilde{\mathbb{M}}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) - d \\
&= y_{jt} - f^0\left(k_{jt}, l_{jt}, m_{jt}\right) - a \ \ \mathbb{M}^{0\ \ -1}\left(k_{jt}, l_{jt}, m_{jt}\right) - \left(1-a\right) \ \ \mathbb{M}^{0\ \ -1}\left(k_{jt}, l_{jt}, m_{jt}\right) - d \\
&= y_{jt} - f^0\left(k_{jt}, l_{jt}, m_{jt}\right) - \ \ \mathbb{M}^{0\ \ -1}\left(k_{jt}, l_{jt}, m_{jt}\right) - d \\
&= \varepsilon_{jt}^{\Theta^0},
\end{aligned}$$

so it trivially satisfies the moment restriction in (ii).

[52] Notice that this is the same set of alternative functions in Theorem 1, replacing for the fact that $\mathbb{M}^{-1}\left(k_{jt}, l_{jt}, m_{jt}\right) = \phi\left(k_{jt}, l_{jt}, m_{jt}\right) - f\left(k_{jt}, l_{jt}, m_{jt}\right)$.

Third, it follows that

$$\begin{aligned}
\eta_{jt}^{\tilde{\Theta}} + \varepsilon_{jt}^{\tilde{\Theta}} &= y_{jt} - \tilde{f}\left(k_{jt}, l_{jt}, m_{jt}\right) - \tilde{h} \quad \tilde{\mathbb{M}}^{-1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d \\
&= y_{jt} - f^0\left(k_{jt}, l_{jt}, m_{jt}\right) - a \quad \mathbb{M}^{0\ -1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d \\
&\quad - \left(1-a\right) h^0 \quad \mathbb{M}^{0\ -1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d \\
&= \underbrace{y_{jt} - f^0\left(k_{jt}, l_{jt}, m_{jt}\right) - h^0 \quad \mathbb{M}^{0\ -1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d}_{\eta_{jt}^{\Theta^0} + \varepsilon_{jt}^{\Theta^0}} \\
&\quad - a \left[\underbrace{\mathbb{M}^{0\ -1}\left(k_{jt}, l_{jt}, m_{jt}\right) + d - h^0 \quad \mathbb{M}^{0\ -1}\left(k_{jt-1}, l_{jt-1}, m_{jt-1}\right) + d}_{\eta_{jt}^{\Theta^0}} \right] \\
&= \left(1-a\right) \eta_{jt}^{\Theta^0} + \varepsilon_{jt}^{\Theta^0}.
\end{aligned}$$

Since $\varepsilon_{jt}^{\tilde{\Theta}} = \varepsilon_{jt}^{\Theta^0}$, it immediately follows that $E_{G_{\eta_{jt}^{\tilde{\Theta}}, \varepsilon_{jt}^{\tilde{\Theta}} | \Gamma_{jt}}} \quad \varepsilon_{jt}^{\tilde{\Theta}} \mid \Gamma_{jt} \quad = 0$. It also follows that $\eta_{jt}^{\tilde{\Theta}} = \left(1-a\right) \eta_{jt}^{\Theta^0}$. By a simple change of variables we have that

$$\begin{aligned}
E_{G_{\eta_{jt}^{\tilde{\Theta}}, \varepsilon_{jt}^{\tilde{\Theta}} | \Gamma_{jt}}} \quad \eta_{jt}^{\tilde{\Theta}} \mid \Gamma_{jt} \quad &= E_{G_{\eta_{jt}^{\tilde{\Theta}} | \Gamma_{jt}}} \quad \eta_{jt}^{\tilde{\Theta}} \mid \Gamma_{jt} \\
&= E_{G_{\eta_{jt}^{\Theta^0} | \Gamma_{jt}}} \quad \frac{\eta_{jt}^{\tilde{\Theta}}}{\left(1-a\right)} \mid \Gamma_{jt} \\
&= E_{G_{\eta_{jt}^{\Theta^0} | \Gamma_{jt}}} \quad \eta_{jt}^{\Theta^0} \mid \Gamma_{jt} \\
&= 0.
\end{aligned}$$

Hence, our alternative $\tilde{\Theta}$ satisfies the moment restriction in (iii).

Finally we notice that, since $\left(\mathbb{M}^0\right)^{-1}$ is invertible given Assumption 3, $\tilde{\mathbb{M}}^{\ -1} \equiv \left(1-a\right)\left(\mathbb{M}^0\right)^{-1}$ is therefore also invertible and hence satisfies Assumption 3 (i.e., (iv)) as well. Since both $\tilde{\Theta}$ and Θ^0 satisfy requirements (i)-(iv), i.e., both rationalize the data, we conclude that $\Theta^0 = \left(f^0, h^0, \mathbb{M}^0\right)$ is not identified. □

Online Appendix O4: Monte Carlo Simulations

We begin by reminding the reader of the general structure of our simulated data. We consider a panel of (up to) 500 firms over (up to) 50 periods, and repeat the simulations (up to) 500 times. To sim-

plify the problem we abstract away from labor and consider the following Cobb-Douglas production function

$$Y_{jt} = K_{jt}^{\alpha_k} M_{jt}^{\alpha_m} e^{\omega_{jt}+\varepsilon_{jt}},$$

where $\alpha_k = 0.25$, $\alpha_m = 0.65$, and ε_{jt} is measurement error that is distributed $N(0, 0.07)$. ω_{jt} follows an AR(1) process

$$\omega_{jt} = \delta_0 + \delta\omega_{jt-1} + \eta_{jt},$$

where $\delta_0 = 0.2$, $\delta = 0.8$, and $\eta_{jt} \sim N(0, 0.04)$. We select the variances of the errors and the AR(1) parameters to roughly correspond to the estimates from our Chilean and Colombian datasets.

The environment facing the firms is the following. At the beginning of each period, firms choose investment I_{jt} and intermediate inputs M_{jt}. Investment determines the next period's capital stock via the law of motion for capital

$$K_{jt+1} = (1 - \kappa_j) K_{jt} + I_{jt},$$

where $\kappa_j \in \{0.05, 0.075, 0.10, 0.125, 0.15\}$ is the depreciation rate which is distributed uniformly across firms. Depending on the simulation, intermediate inputs may be subject to quadratic adjustment costs of the form

$$C_{jt}^M = 0.5b\frac{(M_{jt} - M_{jt-1})^2}{M_{jt}},$$

where $b \in [0, 1]$ is a parameter that indexes the level of adjustment costs, which we vary in our simulations. The case of $b = 0$ corresponds to intermediate inputs m_{jt} being chosen flexibly in period t.

Firms choose investment and intermediate inputs to maximize expected discounted profits. The problem of the firm, written in recursive form, is thus given by

$$\begin{aligned} V(K_{jt}, M_{jt-1}, \omega_{jt}) &= \max_{I_{jt}, M_{jt}} P_t K_{jt}^{\alpha_k} M_{jt}^{\alpha_m} e^{\omega_{jt}} - P_t^I I_{jt} - \rho_t M_{jt} \\ &\quad -0.5b\frac{(M_{jt} - M_{jt-1})^2}{M_{jt}} + \beta E_t V(K_{jt+1}, M_{jt}, \omega_{jt+1}) \\ s.t. & \\ & K_{jt+1} = (1 - \kappa_j) K_{jt} + I_{jt} \\ & I_{jt} \geq 0, M_{jt} \geq 0 \\ & \omega_{jt+1} = \delta_0 + \delta\omega_{jt} + \eta_{jt+1}. \end{aligned}$$

The price of output P_t and the price of intermediate inputs ρ_t are set to 1. The price of investment P_t^I is set to 8, and there are no other costs to investment. The discount factor is set to 0.985.

In order for our Monte Carlo simulations not to depend on the initial distributions of (k, m, ω), we simulate each firm for a total of 200 periods, dropping the first 150 periods. The initial conditions, k_1, m_0, and ω_1 are drawn from the following distributions: $U(11, 400)$, $U(11, 400)$, and $U(1, 3)$. Since the firm's problem does not have an analytical solution, we solve the problem numerically by value function iteration with an intermediate modified policy iteration with 100 steps, using a multi-linear interpolant for both the value and policy functions.[53]

Monte Carlo 1: Time Series Variation

In this set of Monte Carlo simulations, we evaluate the performance of using time series variation as a source of identification for gross output production functions as described in Section 3. Since we are proceeding under the proxy variable approach, we set adjustment costs in intermediate inputs to zero ($b = 0$). We augment our baseline Monte Carlo setup by introducing variation in relative prices and simulate data based on four different levels of variation in relative prices. The first two levels correspond to what we observe in our Colombian and Chilean datasets, respectively. In addition, we create a version with twice the degree of what we observe for Chile (the largest of the two), and another corresponding to ten times the observed variation. In order to examine the importance of sample size, for each of these values of time series variation, we construct 12 different panel structures: 200 vs. 500 firms and 3, 5, 10, 20, 30, and 50 periods.

We estimate a version of the proxy variable technique applied to gross output, as described in Section 3, using intermediate inputs as the proxy. In order to reduce the potential noise from nonparametric estimation, we impose the true parametric structure of the model in the estimation routine (i.e., Cobb-Douglas and the AR(1)). Figure 1 in the main text reports average elasticity estimates across the 500 simulations. In the first panel, we report estimates of the output elasticity of intermediate inputs for 500 firms and for varying numbers of time periods, averaged across 500 simulations. In the second panel, we do the same but with 200 firms. Each line corresponds to a different level of time series variation in prices. For the levels of time series variation corresponding to what we observe in our data (labeled "Colombia" and "Chile"), the proxy variable estimator performs quite poorly (regardless of the number of firms/periods), substantially overestimating the true elasticity of 0.65, and

[53] See Judd (1998) for details.

in some cases generating estimates exceeding 1. For twice the level of time series variation as what we observe in Chile, we start to see some convergence towards to the truth as the number of periods increases, but even for the largest case of 500 firms/50 periods, the estimator is still substantially biased. It is only when we give ourselves ten times the level of variation in Chile that the estimator starts to significantly improve, although again, only when the panel is sufficiently long.

In order to illustrate the precision of the estimator, in Figure O4.1 we plot the 2.5 and 97.5 percentiles of the Monte Carlo estimates (in addition to the mean) for the largest degree of time series variation and largest number of firms. While the mean estimate does converge towards to truth as the number of periods increases, the distribution of the estimates across simulations is quite dispersed. With 10 periods, the 95% interquantile range covers both 0 and 1. Even with 50 periods of data, the range runs from 0.24 to 0.83, implying fairly noisy estimates.

Monte Carlo 2: Performance of Our Baseline Identification Strategy

In order to evaluate the performance of our proposed identification strategy, we simulate 100 datasets consisting of 500 firms and 30 periods each, setting adjustment costs to zero. In order to highlight that our procedure does not rely on time series variation in prices, we impose that relative prices are constant over time. We examine the performance of our estimator under three different underlying production technologies. We first use the baseline Cobb-Douglas technology: $Y_{jt} = K_{jt}^{\alpha_k} M_{jt}^{\alpha_m} e^{\omega_{jt}+\varepsilon_{jt}}$, with $\alpha_k = 0.25, \alpha_m = 0.65$. In the second set of simulations we employ the following CES technology: $Y_{jt} = \left(0.25K_{jt}^{0.5} + 0.65M_{jt}^{0.5}\right)^{\frac{0.9}{0.5}} e^{\omega_{jt}+\varepsilon_{jt}}$. Finally we simulate data from a translog production function, which in logs is given by $y_{jt} = 0.25k_{jt} + 0.65m_{jt} + 0.015k_{jt}^2 + 0.015m_{jt}^2 - 0.032k_{jt}m_t + \omega_{jt} + \varepsilon_{jt}$.

For each of the specifications, we estimate the production function using our nonparametric procedure described in Sections 5. In Table 1 in the main body, we summarize the estimates of the production function from these simulations. For each simulated dataset, we calculate the mean output elasticity of both capital and intermediate inputs (as well as the sum), as well as the standard deviation and the fraction outside of the range of $(0, 1)$. In the table we report the average of these statistics across each of the simulated datasets, as well as the corresponding standard error (calculated across the 100 simulations).

Across all three technology specifications, our procedure performs very well in recovering the

mean elasticities. For both inputs, the average mean elasticities obtained by our procedure are very close to the true values. The largest difference is the mean capital elasticity for CES, which is 0.2196 compared to a true value of 0.2253. They also have consistently very low standard errors.

Not only is our estimator capable of replicating the true mean elasticities, it also does an excellent job of recovering the heterogeneity in elasticities across firms when it exists (CES and translog) and the absence of such heterogeneity when it does not (Cobb-Douglas). This is reflected in the average standard deviations of elasticities that very closely match the truth. Finally we note that in only one case does our estimator produce elasticity estimates outside of the range of 0 to 1 (the capital elasticity for CES). Even then less than 1% fall outside this range.

Monte Carlo 3: Robustness to Adjustment Costs in Flexible Inputs

In our third set of simulations, we evaluate how well our estimator performs when the first-order condition for intermediate inputs does not hold. We generate 100 Monte Carlo samples for each of 9 values of the adjustment cost parameter b, ranging from zero adjustment costs to very large adjustment costs. In each sample we simulate a panel of 500 firms over 30 periods. For the largest value, $b = 1$, this would imply that firms in our Chilean and Colombian datasets, on average, pay substantial adjustment costs for intermediate inputs of almost 10% of the value of total gross output. 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 this procedure provides consistent estimates under the presence of adjustment costs. We compare these estimates to ones obtained via our nonparametric procedure, which assumes adjustment costs of zero.

We impose the (true) Cobb-Douglas parametric form 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. Given the Cobb-Douglas structure and the AR(1) process for productivity, we have

$$y_{jt} - \alpha_k k_{jt} - \alpha_m m_{jt} - \delta_0 - \delta \left(y_{jt-1} - \alpha_k k_{jt-1} - \alpha_m m_{jt-1} \right) = \eta_{jt} - \delta \varepsilon_{jt-1} + \varepsilon_{jt}.$$

The dynamic panel procedure estimates the parameter vector $(\alpha_k, \alpha_m, \delta_0, \delta)$ by forming moments in the RHS of the equation above. Specifically 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. Not surprisingly, the dynamic panel data method breaks down and becomes very unstable for small values of adjustment costs, as these costs are insufficient to provide identifying variation via the lags. This is reflected both in the large percentile ranges and in the fact that the average estimates bounce around the truth. Our method on the other hand performs very well, as expected. This is the case even though for dynamic panel we impose and exploit the restriction that the true technology is Cobb-Douglas, whereas for our procedure we do not.

As we increase the level of adjustment costs, our nonparametric method experiences a small upward bias relative to the truth and relative to dynamic panel, although in some cases our estimates are quite close to those of dynamic panel. The percentile range for dynamic panel is much larger, however. So while on average dynamic panel performs slightly better for large values of adjustment costs, the uncertainty in the estimates is larger. Overall our procedure performs remarkably well, both compared to the truth and to dynamic panel. This is true even for the largest value of adjustment costs ($b = 1$), the worst case for our estimator and best case for dynamic panel. In this case our average estimated elasticity is 0.688 is less than 4 percentage points larger than the truth and about 2.5 percentage points larger than the dynamic panel estimate.

Monte Carlo 4: Inference

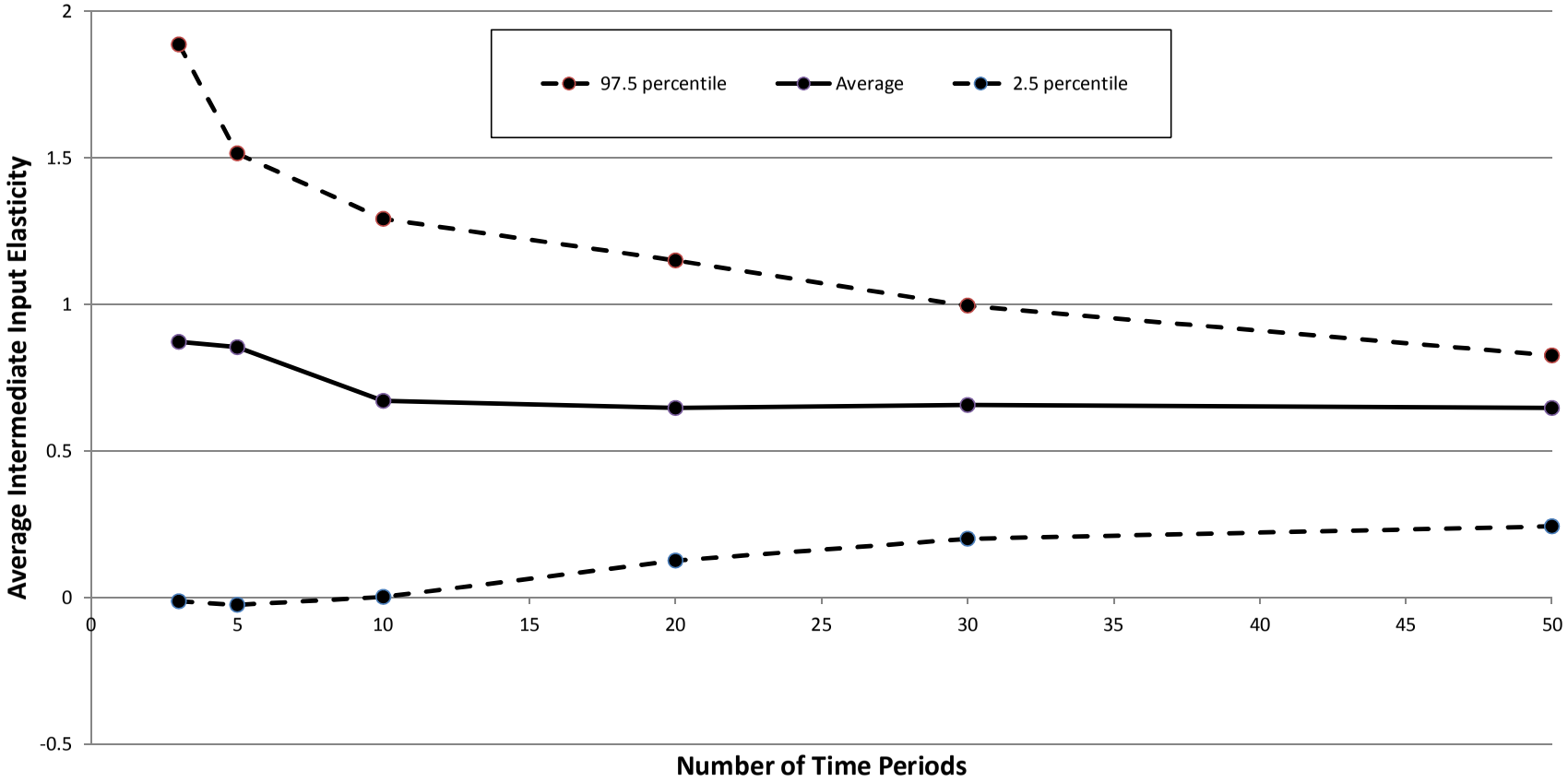
In the final set of Monte Carlo simulations, we provide evidence that our bootstrap procedure has the correct coverage for our estimator. For this set of simulations, we set the adjustment cost parameter for intermediate inputs, b, to zero to correspond with our data generating process. We begin by simulating 500 samples, each consisting of 500 firms over 30 periods. For each sample we nonparametrically bootstrap the data 199 times.[54] For each bootstrap replication we estimate the output elasticities of capital and intermediate inputs using our nonparametric procedure as described in Section 5. We then compute the 95% bootstrap confidence interval using the 199 bootstrap replications. This generates 500 bootstrap confidence intervals, one for each sample. We then count how many times (out of 500) the true values of the output elasticities (0.25 for capital and 0.65 for intermediate inputs) lie within the bootstrap confidence interval. The results are presented graphically in Figures O4.4 and O4.5.

[54] See Davidson and MacKinnon (2004).

The true value of the elasticity is contained inside the 95% confidence interval 95.4% (capital) and 94.2% (intermediate inputs) of the time. Hence, for both the capital and intermediate elasticities, we obtain the correct coverage, suggesting that we can use our bootstrap procedure to do inference even in the nonparametric case.

Figure O4.1: Monte Carlo--Proxy Variable Estimator Applied to Gross Output

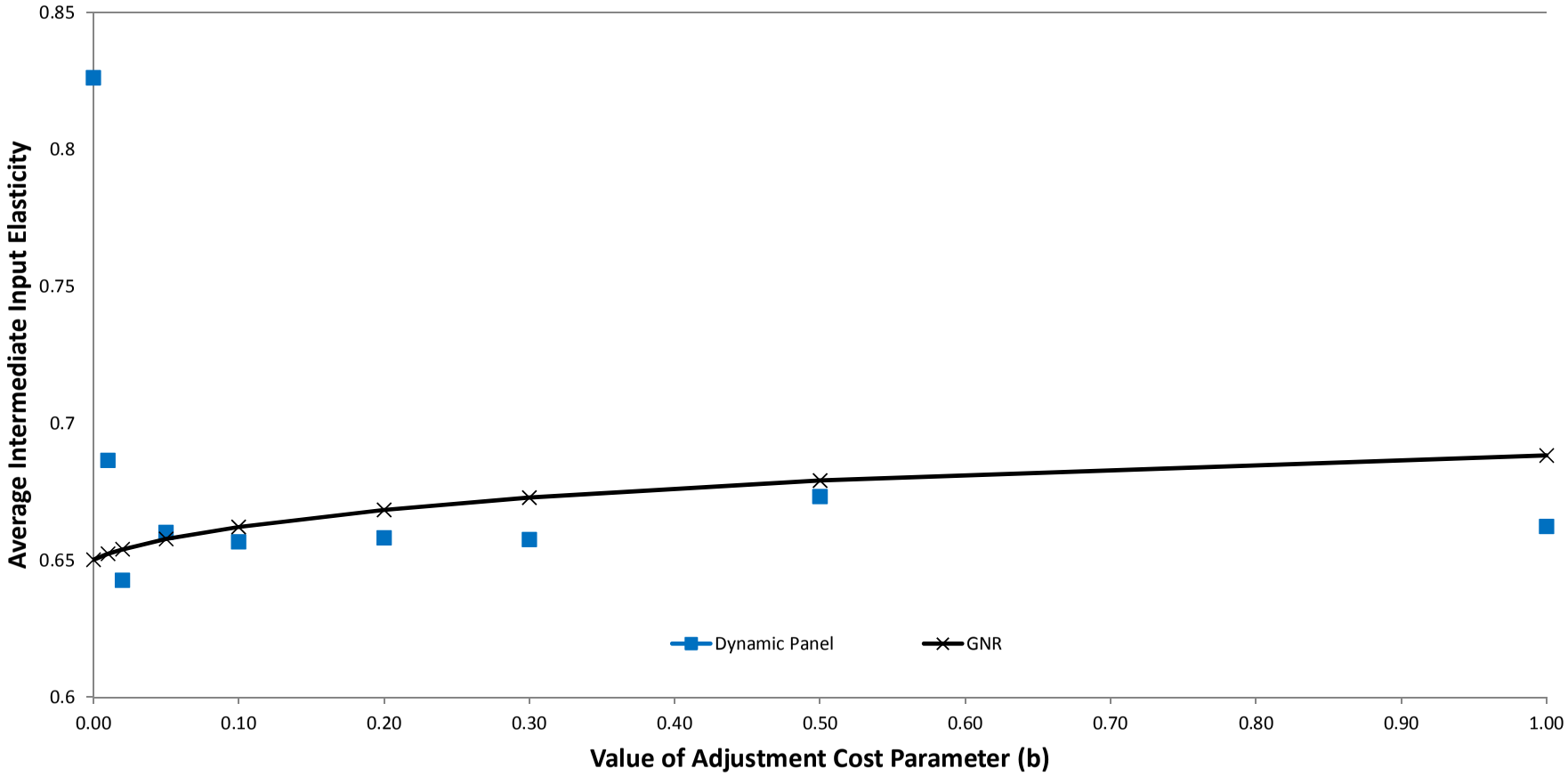
Intermediate Input Elasticity: 500 Firms, Time Series Variation = 10 X Chile



Notes: This figure presents the results from applying a proxy variable estimator extended to gross output to Monte Carlo data generated as described in Online Appendix O4. The data are generated with time-series variation corresponding to ten times the largest amount we observe in our data (Chile). The x-axis measures the number of time periods in the panel used to generate the data. The y-axis measures average and the 2.5 and 97.5 percentiles of the estimated intermediate input elasticity across 100 Monte Carlo simulations. The true value of the elasticity is 0.65.

Figure O4.2: Monte Carlo--Estimator Performance--Intermediate Input Elasticity

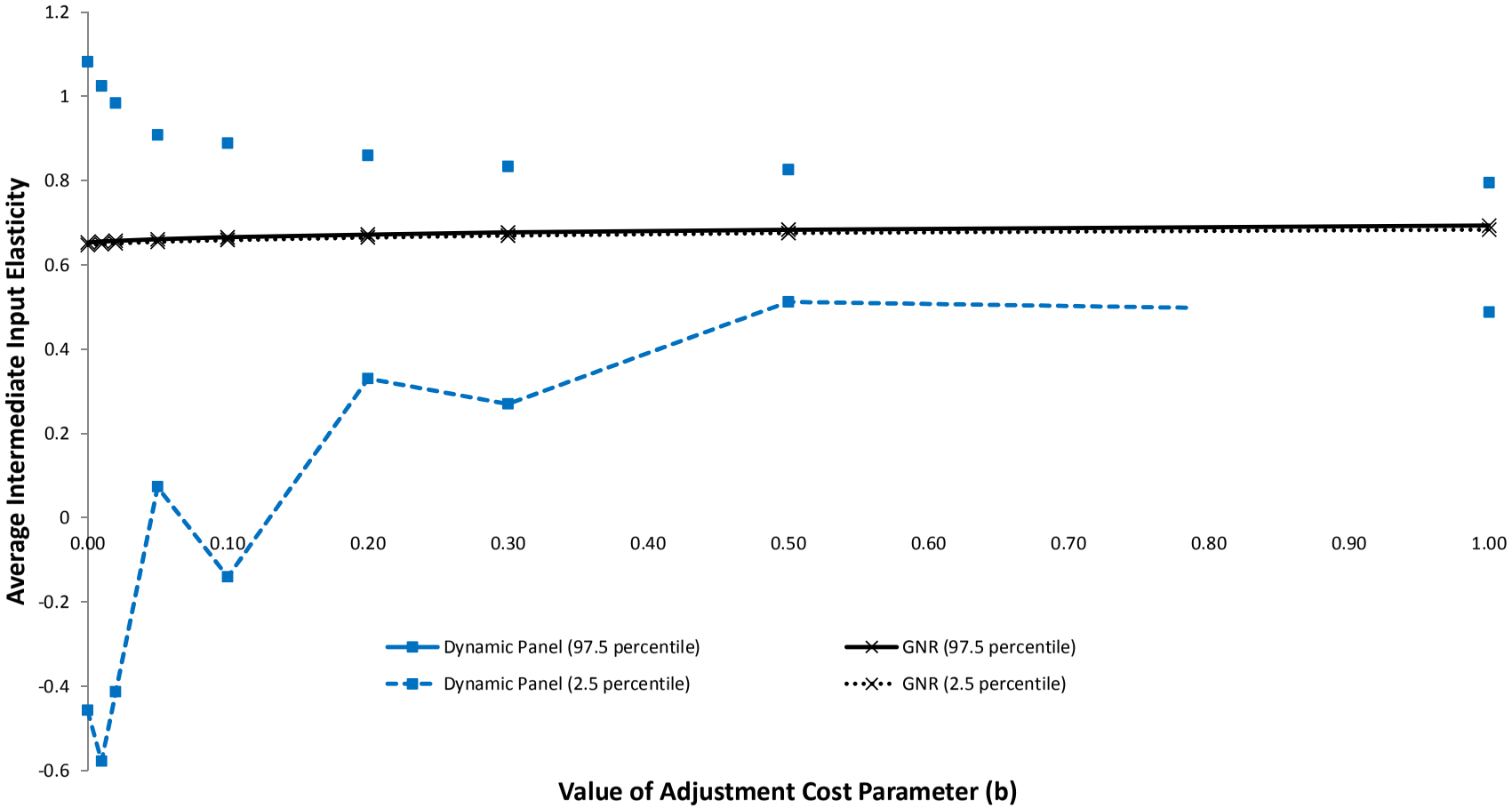
GNR and Dynamic Panel: Averages



Notes: This figure presents the results from applying both our estimator (GNR) and a dynamic panel data estimator to Monte Carlo data generated as described in Online Appendix O4. The data are generated such that the first-order condition no longer holds because of quadratic adjustment costs. The parameter b indexes the degree of adjustment costs in intermediate inputs facing the firm, with higher values representing larger adjustment costs. The y-axis measures the average estimated intermediate input elasticity for both estimators across 100 Monte Carlo simulations. The true value of the average elasticity is 0.65.

Figure O4.3: Monte Carlo--Estimator Performance--Intermediate Input Elasticity

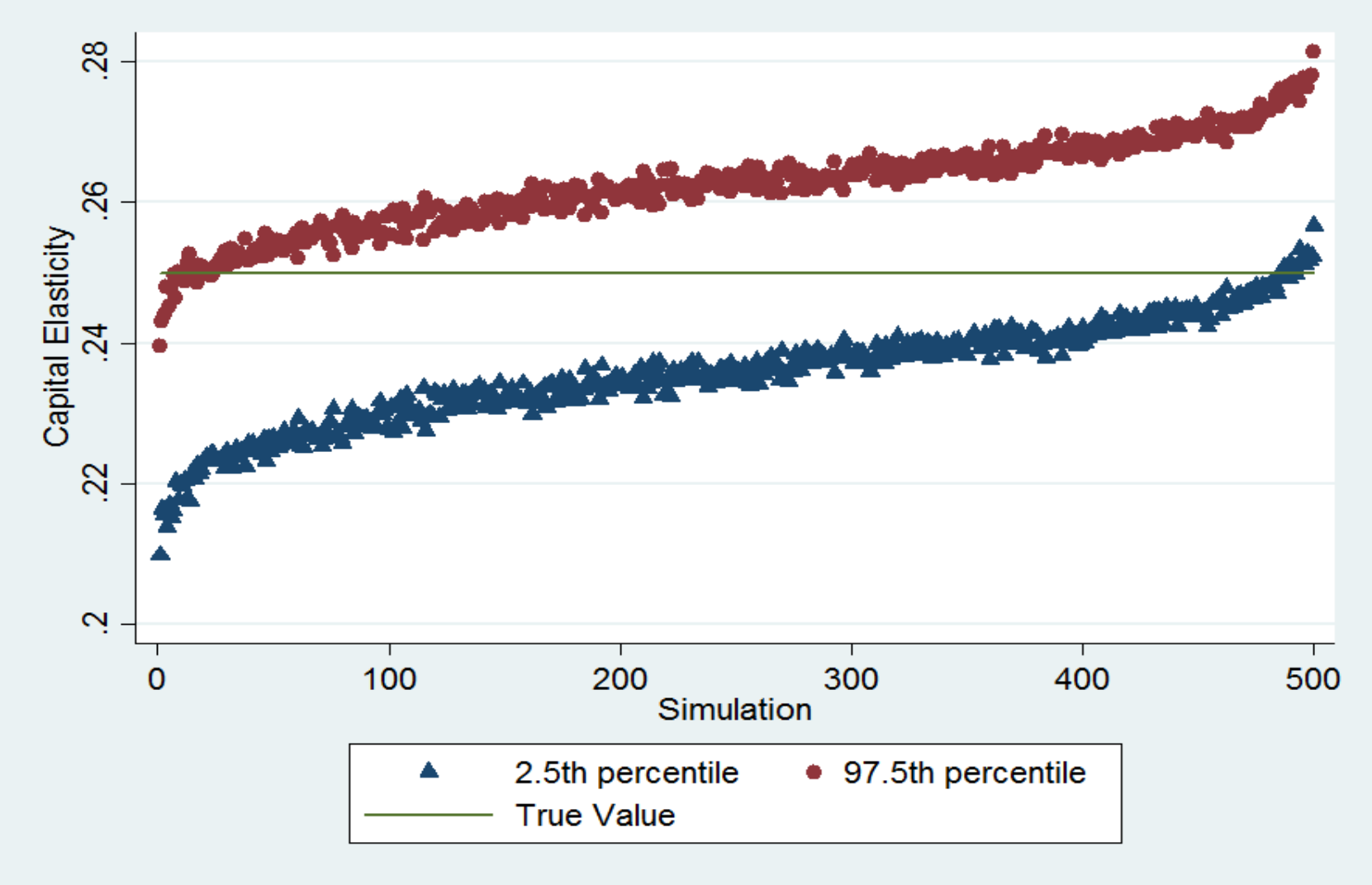
GNR and Dynamic Panel: Percentile Ranges



Notes: This figure presents the results from applying both our estimator (GNR) and a dynamic panel data estimator to Monte Carlo data generated as described in Online Appendix O4. The data are generated such that the first-order condition no longer holds because of quadratic adjustment costs. The parameter b indexes the degree of adjustment costs in intermediate inputs facing the firm, with higher values representing larger adjustment costs. The y-axis measures the 2.5 and 97.5 percentiles of the estimated intermediate input elasticity for both estimators across 100 Monte Carlo simulations. The true value of the average elasticity is 0.65.

Figure O4.4: Monte Carlo--Inference--Capital Elasticity

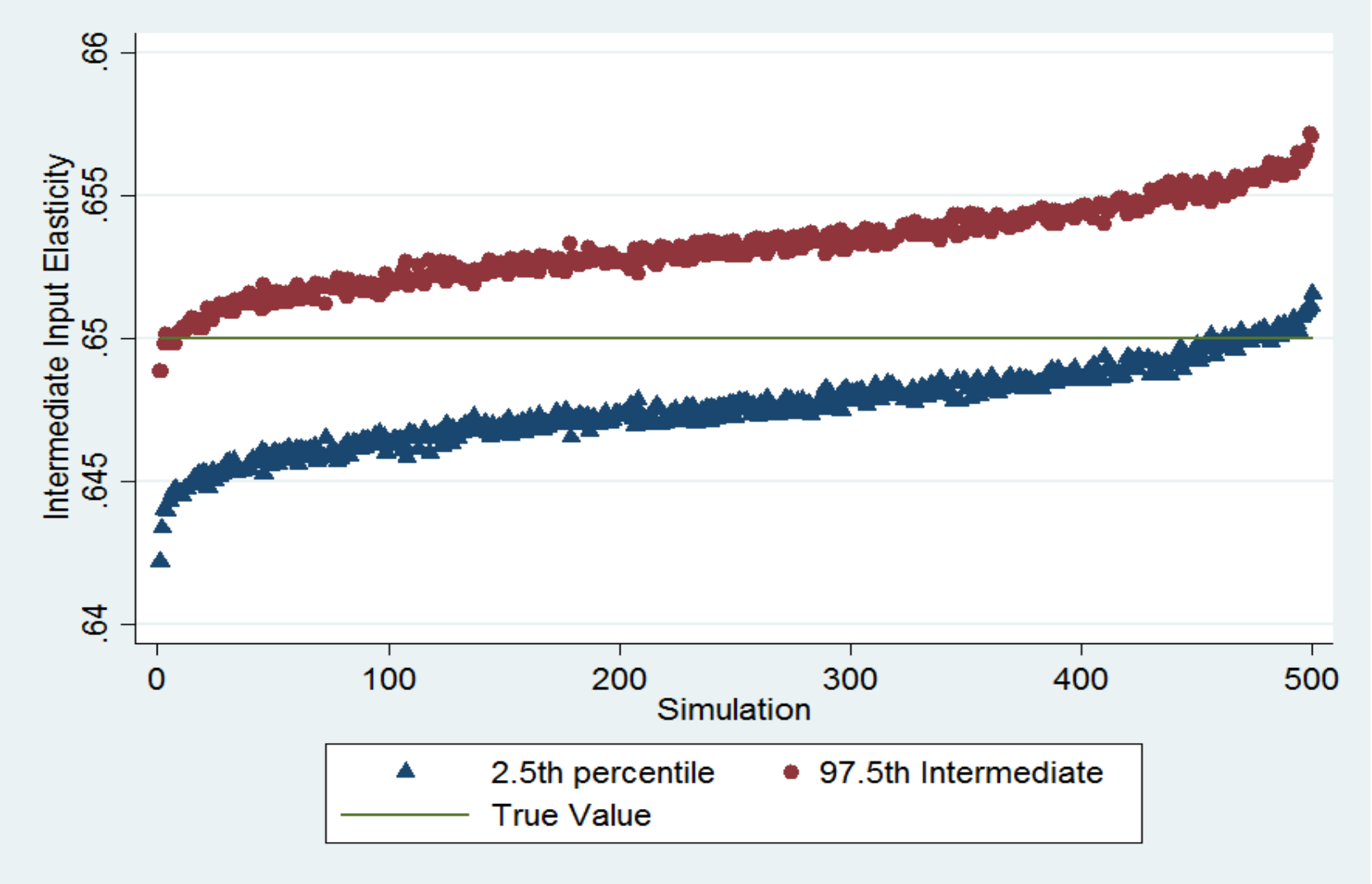
Distribution of 95% Bootstrap Confidence Intervals



Notes: This figure presents the results from applying our estimator to Monte Carlo data generated as described in Online Appendix O4, in the absence of adjustment costs. For each simulation we nonparametrically bootstrap the data 199 times. For each bootstrap replication we estimate the output elasticity of capital using our procedure as described in Section 5. We then compute the 95% bootstrap confidence intervals using these replications. This generates a confidence interval for each of the 500 Monte Carlo samples. In the figure we plot the lower and upper boundaries of the confidence intervals for each Monte Carlo sample. The simulations are sorted by the mid-point of these intervals. The true value of the elasticity is 0.25. 95.4% of the constructed confidence intervals cover the true value.

Figure O4.5: Monte Carlo: Inference--Intermediate Input Elasticity

Distribution of 95% Bootstrap Confidence Intervals



Notes: This figure presents the results from applying our estimator to Monte Carlo data generated as described in Online Appendix O4, in the absence of adjustment costs. For each simulation we nonparametrically bootstrap the data 199 times. For each bootstrap replication we estimate the output elasticity of intermediate inputs using our procedure as described in Section 5. We then compute the 95% bootstrap confidence intervals using these replications. This generates a confidence interval for each of the 500 Monte Carlo samples. In the figure we plot the lower and upper boundaries of the confidence intervals for each Monte Carlo sample. The simulations are sorted by the mid-point of these intervals. The true value of the elasticity is 0.65. 94.2% of the constructed confidence intervals cover the true value.

Online Appendix O5: Extensions

In this section we discuss four modifications to our baseline model: allowing for fixed effects, incorporating additional unobservables in the flexible input demand, allowing for multiple flexible inputs, and revenue production functions.

O5-1. Fixed Effects

One benefit of our identification strategy is that it can easily incorporate fixed effects in the production function. With fixed effects, the production function can be written as

$$y_{jt} = f\left(k_{jt}, l_{jt}, m_{jt}\right) + a_j + \omega_{jt} + \varepsilon_{jt}, \tag{33}$$

where a_j is a firm-level fixed effect.[55] From the firm's perspective, the optimal decision problem for intermediate inputs is the same as before, as is the derivation of the nonparametric share regression (equation (11)), with $\widetilde{\omega}_{jt} \equiv a_j + \omega_{jt}$ replacing ω_{jt}.

The other half of our approach can be easily augmented to allow for the fixed effects. We follow the dynamic panel data literature and impose that persistent productivity ω follows a first-order linear Markov process to difference out the fixed effects: $\omega_{jt} = \delta\omega_{jt-1} + \eta_{jt}$.[56] The equivalent of equation (17) is given by:

$$\mathcal{Y}_{jt} = a_j - \mathscr{C}\left(k_{jt}, l_{jt}\right) + \delta\left(\mathcal{Y}_{jt-1} + \mathscr{C}\left(k_{jt-1}, l_{jt-1}\right)\right) + \eta_{jt}.$$

Subtracting the counterpart for period $t-1$ eliminates the fixed effect. Re-arranging terms leads to:

$$\begin{aligned}\mathcal{Y}_{jt} - \mathcal{Y}_{jt-1} \;=\; & -\left(\mathscr{C}\left(k_{jt}, l_{jt}\right) - \mathscr{C}\left(k_{jt-1}, l_{jt-1}\right)\right) + \delta\left(\mathcal{Y}_{jt-1} - \mathcal{Y}_{jt-2}\right) \\ & +\delta\left(\mathscr{C}\left(k_{jt-1}, l_{jt-1}\right) - \mathscr{C}\left(k_{jt-2}, l_{jt-2}\right)\right) + \left(\eta_{jt} - \eta_{jt-1}\right).\end{aligned}$$

Recall that $E\left[\eta_{jt} \mid \Gamma_{jt}\right] = 0$. Since $\Gamma_{jt-1} \subset \Gamma_{jt}$, this implies that $E\left[\eta_{jt} - \eta_{jt-1} \mid \Gamma_{jt-1}\right] = 0$, where Γ_{jt-1} includes $(k_{jt-1}, l_{jt-1}, \mathcal{Y}_{jt-2}, k_{jt-2}, l_{jt-2}, \mathcal{Y}_{jt-3}, ...)$.

[55] Kasahara et al. (2015) generalize our approach to allow for the entire production function to be firm-specific.

[56] For simplicity we use an AR(1) here, but higher order linear auto-regressive models (e.g., an AR(2)) can be incorporated as well. We omit the constant from the Markov process since it is not separately identified from the mean of the fixed effects.

Let

$$\begin{aligned}\mu\left(k_{jt},l_{jt},k_{jt-1},l_{jt-1},\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right),k_{jt-2},l_{jt-2}\right) &= -\left(\mathscr{C}\left(k_{jt},l_{jt}\right)-\mathscr{C}\left(k_{jt-1},l_{jt-1}\right)\right) \\ &\quad +\delta\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right) \\ &\quad +\delta\left(\mathscr{C}\left(k_{jt-1},l_{jt-1}\right)-\mathscr{C}\left(k_{jt-2},l_{jt-2}\right)\right).\end{aligned} \tag{34}$$

From this we have the following nonparametric IV equation

$$\begin{aligned}&E\left[\mathcal{Y}_{jt}-\mathcal{Y}_{jt-1}\mid k_{jt-1},l_{jt-1},\mathcal{Y}_{jt-2},k_{jt-2},l_{jt-2},k_{jt-3},l_{jt-3}\right] \\ &=E\left[\mu\left(k_{jt},l_{jt},k_{jt-1},l_{jt-1},\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right),k_{jt-2},l_{jt-2}\right)\mid k_{jt-1},l_{jt-1},\mathcal{Y}_{jt-2},k_{jt-2},l_{jt-2},k_{jt-3},l_{jt-3}\right],\end{aligned}$$

which is an analogue to equation (19) in the case without fixed effects.

Theorem 7. *Under Assumptions 2 - 4, plus the additional assumptions of an AR(1) process for ω and that the distribution of the endogenous variables conditional on the exogenous variables (i.e., instruments),*
$G\left(k_{jt},l_{jt},k_{jt-1},l_{jt-1},\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right),k_{jt-2},l_{jt-2}\mid k_{jt-3},l_{jt-3},k_{jt-1},l_{jt-1},\mathcal{Y}_{jt-2},k_{jt-2},l_{jt-2}\right)$, is complete (as defined in Newey and Powell, 2003), the production function f is nonparametrically identified up to an additive constant if $\frac{\partial}{\partial m_{jt}}f\left(k_{jt},l_{jt},m_{jt}\right)$ is nonparametrically known.

Following the first part of the proof of Theorem 3, we know that the production function is identified up to an additive function $\mathscr{C}\left(k_{jt},l_{jt}\right)$. Following directly from Newey and Powell (2003), we know that, if the distribution G is complete, then the function $\mu\left(\right)$ defined in equation (34) is identified.

Let $\tilde{\mathscr{C}},\tilde{\delta}$ be a candidate alternative pair of functions. $\left(\mathscr{C},\delta\right)$ and $\tilde{\mathscr{C}},\tilde{\delta}$ are observationally equivalent if and only if

$$\begin{aligned}&-\left(\mathscr{C}\left(k_{jt},l_{jt}\right)-\mathscr{C}\left(k_{jt-1},l_{jt-1}\right)\right)+\delta\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right)+\delta\left(\mathscr{C}\left(k_{jt-1},l_{jt-1}\right)-\mathscr{C}\left(k_{jt-2},l_{jt-2}\right)\right) \\ &=-\ \tilde{\mathscr{C}}\left(k_{jt},l_{jt}\right)-\tilde{\mathscr{C}}\left(k_{jt-1},l_{jt-1}\right)\ +\tilde{\delta}\left(\mathcal{Y}_{jt-1}-\mathcal{Y}_{jt-2}\right)+\tilde{\delta}\ \tilde{\mathscr{C}}\left(k_{jt-1},l_{jt-1}\right)-\tilde{\mathscr{C}}\left(k_{jt-2},l_{jt-2}\right)\ .\end{aligned} \tag{35}$$

By taking partial derivatives of both sides of (35) with respect to k_{jt} and l_{jt} we obtain

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

for $z \in \{k_{jt}, l_{jt}\}$, which implies $\mathscr{C}(k_{jt}, l_{jt}) - \tilde{\mathscr{C}}(k_{jt}, l_{jt}) = c$ for a constant c. Thus we have shown the production function is identified up to an additive constant.

The estimation strategy for the model with fixed effects is almost exactly the same as without fixed effects. The first stage, estimating $\mathcal{D}_r(k_{jt}, l_{jt}, m_{jt})$, is the same. We then form $\hat{\mathcal{Y}}_{jt}$ in the same way. We also use the same series estimator for $\mathscr{C}(k_{jt}, l_{jt})$. This generates an analogue to equation (24):

$$\begin{aligned} \mathcal{Y}_{jt} - \mathcal{Y}_{jt-1} &= - \sum_{0<\tau_k+\tau_l\leq\tau} \alpha_{\tau_k,\tau_l} k_{jt}^{\tau_k} l_{jt}^{\tau_l} + \delta\left(\mathcal{Y}_{jt-1} - \mathcal{Y}_{jt-2}\right) \\ &\quad + (\delta+1) \sum_{0<\tau_k+\tau_l\leq\tau} \alpha_{\tau_k,\tau_l} k_{jt-1}^{\tau_k} l_{jt-1}^{\tau_l} \\ &\quad - \delta \sum_{0<\tau_k+\tau_l\leq\tau} \alpha_{\tau_k,\tau_l} k_{jt-2}^{\tau_k} l_{jt-2}^{\tau_l} + (\eta_{jt} - \eta_{jt-1}). \end{aligned} \tag{36}$$

We can use similar moments as for the model without fixed effects, except that now we need to lag the instruments one period given the differencing involved. Therefore the following moments can be used to form a standard sieve GMM criterion function to estimate (α, δ): $E\left[(\eta_{jt} - \eta_{jt-1}) k_{jt-\iota}^{\tau_k} l_{jt-\iota}^{\tau_l}\right]$, for $\iota \geq 1$.

O5-2. Extra Unobservables

Our identification and estimation approach can also be extended to incorporate additional unobservables driving the intermediate input demand. In our baseline model, our system of equations consists of the share equation and the production function given by

$$\begin{aligned} s_{jt} &= \ln D(k_{jt}, l_{jt}, m_{jt}) + \ln \mathcal{E} - \varepsilon_{jt} \\ y_{jt} &= f(k_{jt}, l_{jt}, m_{jt}) + \omega_{jt} + \varepsilon_{jt}. \end{aligned}$$

We now show that our model can be extended to include an additional structural unobservable to the share equation for intermediate inputs, which we denote by ψ_{jt}:

$$\begin{aligned} s_{jt} &= \ln D(k_{jt}, l_{jt}, m_{jt}) + \ln \widetilde{\mathcal{E}} - \varepsilon_{jt} - \psi_{jt} \\ y_{jt} &= f(k_{jt}, l_{jt}, m_{jt}) + \omega_{jt} + \varepsilon_{jt}, \end{aligned} \tag{37}$$

where $\widetilde{\mathcal{E}} \equiv E\left[e^{\psi_{jt}+\varepsilon_{jt}}\right]$.

Assumption 7. *$\psi_{jt} \in \mathcal{I}_{jt}$ is known to the firm at the time of making its period t decisions and is not*

persistent: $P_\psi\left(\psi_{jt} \mid \mathcal{I}_{jt-1}\right) = P_\psi\left(\psi_{jt}\right)$.

O5-2.1. Interpretations for the extra unobservable

We now discuss some possible interpretations for the non-persistent extra unobservable ψ, arising from potentially persistent shocks to the firm's problem.

Shocks to prices of output and/or intermediate inputs Suppose that the prices of output and intermediate inputs, P_t and ρ_t, are not fully known when firm j decides its level of intermediate inputs, but that the firm has private signals about the prices, denoted P_{jt}^* and ρ_{jt}^*, where[57]

$$\begin{aligned} \ln P_{jt}^* &= \ln P_t - \xi_{jt}, \\ \ln \rho_{jt}^* &= \ln \rho_t - \xi_{jt}^M. \end{aligned}$$

Notice that, ex-post, once production occurs and profits are realized, firms can infer the true prices. As a consequence, ξ_{jt-1}, ξ_{jt-1}^M are in the firm's information set in period t, $\mathcal{I}_{jt}$. We allow the noise in the signals ξ, ξ^M to be potentially serially correlated by writing them as

$$\begin{aligned} \xi_{jt} &= g\left(\xi_{jt-1}\right) + \nu_{jt}, \\ \xi_{jt}^M &= g^M \ \xi_{jt-1}^M \ + \nu_{jt}^M. \end{aligned}$$

For concreteness we assume a first-order Markov process, but other processes can be accommodated as long as they can be expressed as functions of $\mathcal{I}_{jt}$ and a separable innovation.

Firms maximize expected profits conditional on their signals:

$$\begin{aligned} \mathbb{M}\left(k_{jt}, l_{jt}, \omega_{jt}\right) &= \max_{M_{jt}} E_{\varepsilon,\nu,\nu^M} \ P_t F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}+\varepsilon_{jt}} - \rho_t M_{jt} \mid \mathcal{I}_{jt} \\ &= \max_{M_{jt}} E_{\varepsilon,\nu,\nu^M} \quad P_{jt}^* e^{g(\xi_{jt-1})} e^{\nu_{jt}} \ F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}+\varepsilon_{jt}} - \ \rho_{jt}^* e^{g^M\left(\xi_{jt-1}^M\right)} e^{\nu_{jt}^M} \ M_{jt} \mid \mathcal{I}_{jt} \\ &\quad \max_{M_{jt}} E\left(e^{\nu_{jt}}\right) E\left(e^{\varepsilon_{jt}}\right) P_{jt}^* e^{g(\xi_{jt-1})} F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\omega_{jt}} - E \ e^{\nu_{jt}^M} \ \rho_{jt}^* e^{g^M\left(\xi_{jt-1}^M\right)} M_{jt}. \end{aligned}$$

[57] Since only relative prices matter, we could alternatively rewrite the problem in terms of a single signal about relative prices.

This implies that the firm's first-order condition for intermediate inputs is given by

$$E\left(e^{\nu_{jt}}\right)E\left(e^{\varepsilon_{jt}}\right)P_{jt}^{*}e^{g(\xi_{jt-1})}\frac{\partial}{\partial M_{jt}}F\left(k_{jt},l_{jt},m_{jt}\right)e^{\omega_{jt}}-E\quad e^{\nu_{jt}^{M}}\quad\rho_{jt}^{*}e^{g^{M}\left(\xi_{jt-1}^{M}\right)}=0,$$

which can be rewritten as

$$\left(\frac{\rho_{t}M_{jt}}{P_{t}Y_{jt}}\right)=\frac{E\left(e^{\nu_{jt}}\right)E\left(e^{\varepsilon_{jt}}\right)}{E\quad e^{\nu_{jt}^{M}}}\frac{\partial}{\partial m_{jt}}f\left(k_{jt},l_{jt},m_{jt}\right)\frac{e^{\nu_{jt}^{M}}}{e^{v_{jt}}e^{\varepsilon_{jt}}}.$$

Letting $\psi_{jt}\equiv\nu_{jt}-\nu_{jt}^{M}$, we have

$$\ln\frac{\rho_{t}M_{jt}}{P_{t}Y_{jt}}=s_{jt}=\ln D\left(k_{jt},l_{jt},m_{jt}\right)+\ln\widetilde{\mathcal{E}}-\varepsilon_{jt}-\psi_{jt}.$$

Optimization error Suppose that firms do not exactly know their productivity, ω_{jt}, when they make their intermediate input decision. Instead, they observe a signal about productivity $\omega_{jt}^{*}=\omega_{jt}-\xi_{jt}$, where ξ_{jt} denotes the noise in the signal, and similarly to above

$$\xi_{jt}=g\left(\xi_{jt-1}\right)+\psi_{jt}.$$

Ex-post, once production occurs, the firm can infer the true ω. As a consequence, ξ_{jt-1} is in the firm's information set in period t, $\mathcal{I}_{jt}$.

The firm's profit maximization problem with respect to intermediate inputs is

$$\mathbb{M}\left(k_{jt},l_{jt},\omega_{jt}\right)\quad=\quad\arg\max_{M_{jt}}P_{t}E_{\varepsilon,\psi}\quad F\left(k_{jt},l_{jt},m_{jt}\right)e^{\omega_{jt}^{*}+\psi_{jt}+\varepsilon_{jt}}\quad-\rho_{t}M_{jt}.$$

This implies the following first-order condition

$$P_{t}\frac{\partial}{\partial M_{jt}}F\left(k_{jt},l_{jt},m_{jt}\right)e^{\omega_{jt}^{*}}e^{g(\xi_{jt-1})}E_{\varepsilon,\psi}\quad e^{\psi_{jt}+\varepsilon_{jt}}\quad=\rho_{t}$$

Re-arranging to solve for the share of intermediate inputs gives us the share equation

$$\ln\frac{\rho_{t}M_{jt}}{P_{t}Y_{jt}}=s_{jt}=\ln D\left(k_{jt},l_{jt},m_{jt}\right)+\ln\widetilde{\mathcal{E}}-\varepsilon_{jt}-\psi_{jt}.$$

Notice that for both interpretations of ψ, the firm will take into account the value of $\widetilde{\mathcal{E}}\equiv E\quad e^{\varepsilon_{jt}+\psi_{jt}}$

when deciding on the level of intermediate inputs, which means we want to correct the share estimates by this term. As in the baseline model, we can recover this term by estimating the share equation, forming the residuals, $\varepsilon_{jt} + \psi_{jt}$, and computing the expectation of $e^{\varepsilon_{jt}+\psi_{jt}}$.

O5-2.2. Identification

The identification of the share equation is similar to our main specification, but with two differences. The first is that, since ψ_{jt} drives intermediate input decisions and is in the residual of the modified share equation (37), intermediate inputs are now endogenous in the share equation. As a result, we need to instrument for m_{jt} in the share regression. We can use m_{jt-1} as an instrument for m_{jt}, since it is correlated with m_{jt} and independent of the error $(\varepsilon_{jt} + \psi_{jt})$. Since in the share regression we condition only on k_{jt} and l_{jt} (and no lags), m_{jt-1} generates variation in m_{jt} (conditional on k_{jt} and l_{jt}), due to Assumptions 3 and 5. Identification follows from standard nonparametric IV arguments as in Newey and Powell (2003).

The second difference is that the error in the share equation is $\varepsilon_{jt} + \psi_{jt}$ instead of ε_{jt}. We can form an alternative version of $\mathcal{Y}_{jt}$, which we denote $\widetilde{\mathcal{Y}_{jt}}$:

$$\widetilde{\mathcal{Y}_{jt}} \equiv y_{jt} - \quad D\left(k_{jt}, l_{jt}, m_{jt}\right) dm_{jt} - \left(\varepsilon_{jt} + \psi_{jt}\right) = \mathcal{Y}_{jt} - \psi_{jt}. \tag{38}$$

This generates an analogous equation to equation (16) in the paper:

$$\widetilde{\mathcal{Y}_{jt}} = -\mathscr{C}\left(k_{jt}, l_{jt}\right) + \omega_{jt} - \psi_{jt} \Rightarrow \omega_{jt} = \widetilde{\mathcal{Y}_{jt}} + \mathscr{C}\left(k_{jt}, l_{jt}\right) + \psi_{jt}.$$

Re-arranging terms and plugging in the Markovian structure of ω gives us:

$$\widetilde{\mathcal{Y}_{jt}} = -\mathscr{C}\left(k_{jt}, l_{jt}\right) + h\left(\underbrace{\widetilde{\mathcal{Y}_{jt-1}} + \mathscr{C}\left(k_{jt-1}, l_{jt-1}\right) + \psi_{jt-1}}_{\omega_{jt-1}}\right) + \eta_{jt} - \psi_{jt}, \tag{39}$$

which is an analogue of equation (17).

The challenge is that we cannot form ω_{jt-1}, the argument of h in equation (39), because ψ_{jt-1} is not observed. We can, however, construct two noisy measures of ω_{jt-1}: $(\omega_{jt-1} + \varepsilon_{jt-1})$ and

$(\omega_{jt-1} - \psi_{jt-1})$ where

$$\begin{aligned}\omega_{jt-1} + \varepsilon_{jt-1} &= y_{jt-1} - f(k_{jt-1}, l_{jt-1}, m_{jt-1}) \\ &= y_{jt-1} - \quad D(k_{jt-1}, l_{jt-1}, m_{jt-1})\, dm_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1})\end{aligned}$$

$$\begin{aligned}\omega_{jt-1} - \psi_{jt-1} &= (\omega_{jt-1} + \varepsilon_{jt-1}) - (\varepsilon_{jt-1} + \psi_{jt-1}) \\ &= \left(y_{jt-1} - \quad D(k_{jt-1}, l_{jt-1}, m_{jt-1})\, dm_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1})\right) \\ &\quad - (s_{jt-1} - \ln D(k_{jt-1}, l_{jt-1}, m_{jt-1}))\,.\end{aligned}$$

We could proceed to identify h and $\mathscr{C}$ from equation (39) by adopting methods from the measurement error literature (Hu and Schennach, 2008 and Cunha et al., 2010) using one of the noisy measures as our measure of ω_{jt-1} and using the other as an instrument. However, such an exercise is not straightforward and is beyond the scope of the current paper.

Instead, we illustrate our approach using an AR(1) process for the evolution of ω: $h(\omega_{jt-1}) = \delta_0 + \delta\omega_{jt-1} + \eta_{jt}$. We can then re-write equation (39) as

$$\widetilde{\mathcal{Y}_{jt}} = -\mathscr{C}(k_{jt}, l_{jt}) + \delta_0 + \delta \quad \mathcal{Y}_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1}) \quad + \eta_{jt} - \psi_{jt} + \delta\psi_{jt-1}, \tag{40}$$

where now the residual is given by $\eta_{jt} - \psi_{jt} + \delta\psi_{jt-1}$. Given Assumptions 2 and 7, we have that $E[\eta_{jt} - \psi_{jt} + \delta\psi_{jt-1} \mid \Gamma_{jt-1}] = 0$, where recall that $\Gamma_{jt-1} = \Gamma(\mathcal{I}_{jt-2})$, i.e., a transformation of the period $t-2$ information set. If we let

$$\mu \quad k_{jt}, l_{jt}, \mathcal{Y}_{jt-1}, k_{jt-1}, l_{jt-1} \quad = -\mathscr{C}(k_{jt}, l_{jt}) + \delta_0 + \delta \quad \mathcal{Y}_{jt-1} + \mathscr{C}(k_{jt-1}, l_{jt-1}) \quad ,$$

then identification of equation (40) follows from a parallel argument to that in Theorem 7 (i.e., including the completeness assumption and following the nonparametric IV identification arguments in Newey and Powell, 2003). Therefore we can identify the entire production function up to an additive constant. We can also identify δ_0 and δ, as well as productivity: $\omega + \varepsilon$.

O5-3. Multiple Flexible Inputs

Suppose that, in addition to intermediate inputs being flexible, the researcher believes that one or more additional inputs are also flexible.[58] Our approach can also be extended to handle this case. In what follows we assume that labor is the additional flexible input, but the approach can be extended to allow for more than two flexible inputs.

When labor and intermediate inputs are both assumed to be flexible, we have two share equations. We use superscripts M and L to distinguish them. The system of equations is then given by

$$\begin{aligned} s_{jt}^{M} &= \ln D^{M}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ln \mathcal{E} - \varepsilon_{jt} \\ s_{jt}^{L} &= \ln D^{L}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ln \mathcal{E} - \varepsilon_{jt} \\ y_{jt} &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \omega_{jt} + \varepsilon_{jt}. \end{aligned} \tag{41}$$

These two input elasticities define a system of partial differential equations of the production function. By the fundamental theorem of calculus we have

$$\int_{m_0}^{m_{jt}} \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) dm_{jt} = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathscr{C}^{M}\left(k_{jt}, l_{jt}\right)$$

and

$$\int_{l_0}^{l_{jt}} \frac{\partial}{\partial l_{jt}} f\left(k_{jt}, l_{jt}, m_{jt}\right) dl_{jt} = f\left(k_{jt}, l_{jt}, m_{jt}\right) + \mathscr{C}^{L}\left(k_{jt}, m_{jt}\right)$$

where now we have two constants of integration, one for each integrated share equation, $\mathscr{C}^{M}\left(k_{jt}, l_{jt}\right)$ and $\mathscr{C}^{L}\left(k_{jt}, m_{jt}\right)$. Following directly from Varian (1992), these partial differential equations can be combined to construct the production function as follows:

$$f\left(k_{jt}, l_{jt}, m_{jt}\right) = \int_{m_0}^{m_{jt}} \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_0, s\right) ds + \int_{l_0}^{l_{jt}} \frac{\partial}{\partial l_{jt}} f\left(k_{jt}, \tau, m_{jt}\right) d\tau - \mathscr{C}\left(k_{jt}\right). \tag{42}$$

That is, by integrating the (log) elasticities of intermediate inputs and labor, we can construct the

[58] See, for example, Doraszelski and Jaumandreu (2013, 2015).

production function up to a constant that is a function of capital only.[59] Identification of $\mathscr{C}$ and h can be achieved in the same way as described in Section 4 for equation (18), with the difference that in this case $\mathscr{C}$ only depends on capital.

Notice that the model described by equation (41) imposes the testable restriction that the residuals in both share equations are equivalent. If this restriction does not hold, then we could allow for an additional structural error in the model, ψ, as described in the preceding sub-section.[60,61] Our system of equations is thus given by:

$$\begin{aligned} s_{jt}^{M} &= \ln D^{M}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ln \widetilde{\mathcal{E}}^{M} - \varepsilon_{jt} - \psi_{jt}^{M} \\ s_{jt}^{L} &= \ln D^{L}\left(k_{jt}, l_{jt}, m_{jt}\right) + \ln \widetilde{\mathcal{E}}^{L} - \varepsilon_{jt} - \psi_{jt}^{L} \\ y_{jt} &= f\left(k_{jt}, l_{jt}, m_{jt}\right) + \omega_{jt} + \varepsilon_{jt}. \end{aligned}$$

Nonparametric identification of the flexible input elasticities of L and M proceeds as in Appendix O5-2. One can then integrate up the system of partial differential equations as above. Next we can construct an analogue to equation (38) above using the residual from either share equation. Using the intermediate input share equation, we have

$$\widetilde{\widetilde{\mathcal{Y}}}_{jt} \equiv y_{jt} - \int_{m_0}^{m_{jt}} \frac{\partial}{\partial m_{jt}} f\left(k_{jt}, l_0, s\right) ds - \int_{l_0}^{l_{jt}} \frac{\partial}{\partial l_{jt}} f\left(k_{jt}, \tau, m_{jt}\right) d\tau - \varepsilon_{jt} + \psi_{jt}^{M} \tag{43}$$

By subtracting equation (43) from the production function and re-arranging terms we have

$$\widetilde{\widetilde{\mathcal{Y}}}_{jt} = -\mathscr{C}\left(k_{jt}\right) + \omega_{jt} - \psi_{jt}^{M}.$$

Plugging in the Markovian structure of ω gives us

[59] In order to see why this is the case, evaluate the integrals on the RHS of equation (42), we have the following

$$\begin{aligned} f\left(k_{jt}, l_{jt}, m_{jt}\right) &= f\left(k_{jt}, l_0, m_{jt}\right) - \mathscr{C}^{M}\left(k_{jt}, l_0\right) - f\left(k_{jt}, l_0, m_0\right) - \mathscr{C}^{M}\left(k_{jt}, l_0\right) \\ &\quad + f\left(k_{jt}, l_{jt}, m_{jt}\right) - \mathscr{C}^{L}\left(k_{jt}, m_{jt}\right) - f\left(k_{jt}, l_0, m_{jt}\right) - \mathscr{C}^{L}\left(k_{jt}, m_{jt}\right) \\ &\quad + f\left(k_{jt}, l_0, m_0\right) \\ &= f\left(k_{jt}, l_{jt}, m_{jt}\right), \end{aligned}$$

where $f\left(k_{jt}, l_0, m_0\right) \equiv \mathscr{C}\left(k_{jt}\right)$ is a constant of integration that is a function of capital k_{jt}.

[60] In this case since there are two flexible inputs, we allow for two errors, ψ^{M} and ψ^{L}, corresponding to intermediate inputs and labor, respectively. In principle allowing for just one additional error is sufficient, but we add both for symmetry.

[61] Alternatively, it may be possible to allow for other sources of productivity heterogeneity, such as a factor-biased component of technological change as in Doraszelski and Jaumandreu (2015).

$$\widetilde{\widetilde{\mathcal{Y}}}_{jt} = -\mathscr{C}\left(k_{jt}\right) + h\left(\underbrace{\mathcal{Y}_{jt-1} + \mathscr{C}\left(k_{jt-1}\right) + \psi^{M}_{jt-1}}_{\omega_{jt-1}}\right) + \eta_{jt} - \psi^{M}_{jt}, \tag{44}$$

an analogue to equation (39). Identification of $\mathscr{C}$ and h can be achieved in the same way as described in O5-2 for equation (39), with the difference that in this case $\mathscr{C}$ only depends on capital.

O5-4. Revenue Production Functions

We now show that our empirical strategy can be adapted to the setting with imperfect competition and revenue production functions such that 1) we solve the identification problem with flexible inputs and 2) we can recover time-varying industry markups.[62] We specify a generalized version of the demand system in Klette and Griliches (1996) and De Loecker (2011),

$$\frac{P_{jt}}{\Pi_t} = \left(\frac{Y_{jt}}{Y_t}\right)^{\frac{1}{\sigma_t}} e^{\chi_{jt}}, \tag{45}$$

where P_{jt} is the output price of firm j, Π_t is the industry price index, Y_t is a quantity index that plays the role of an aggregate demand shifter,[63] χ_{jt} is an observable (to the firm) demand shock, and σ_t is the elasticity of demand that is allowed to vary over time.

Substituting for price using equation (45), the firm's first-order condition with respect to M_{jt} in the (expected) profit maximization problem is

$$\left(\frac{1}{\sigma_t}+1\right)\Pi_t \frac{Y_{jt}^{\frac{1}{\sigma_t}}}{Y_t^{\frac{1}{\sigma_t}}} \frac{1}{e^{\frac{1}{\sigma_t}\varepsilon_{jt}}} \frac{\partial}{\partial M_{jt}} F\left(k_{jt}, l_{jt}, m_{jt}\right) e^{\chi_{jt}} E\left[e^{\varepsilon_{jt}\left(\frac{1}{\sigma_t}+1\right)}\right] = \rho_t.$$

Following the same strategy as before, we can rewrite this expression in terms of the observed log revenue share, which becomes

$$s_{jt} = \ln\left(\frac{1}{\sigma_t}+1\right) + \ln\left(D\left(k_{jt}, l_{jt}, m_{jt}\right) E\left[e^{\varepsilon_{jt}\left(\frac{1}{\sigma_t}+1\right)}\right]\right) - \left(\frac{1}{\sigma_t}+1\right)\varepsilon_{jt}, \tag{46}$$

[62] This stands in contrast to the Klette and Griliches (1996) approach that can only allow for a markup that is time-invariant.

[63] As noted by Klette and Griliches (1996) and De Loecker (2011), Y_t can be calculated using a market-share weighted average of deflated revenues.

where $s_{jt} \equiv \ln \left[\frac{\rho_t M_{jt}}{P_{jt} Y_{jt}} \right], \frac{1}{\left(\frac{1}{\sigma_t}+1\right)}$ is the expected markup, $D(\cdot)$ is the output elasticity of intermediate inputs, and ε_{jt} is the ex-post shock. Equation (46) nests the one obtained for the perfectly competitive case in (11), the only difference being the addition of the expected markup, which is equal to 1 under perfect competition.

We now show how to use the share regression (46) to identify production functions among imperfectly competitive firms. Letting $\widetilde{\varepsilon}_{jt} = \left[\frac{1}{\sigma_t} + 1 \right] \varepsilon_{jt}$, equation (46) becomes

$$s_{jt} = \Upsilon_t + \ln D(k_{jt}, l_{jt}, m_{jt}) + \ln \widetilde{\mathcal{E}} - \widetilde{\varepsilon}_{jt}, \tag{47}$$

where $\widetilde{\mathcal{E}} = E\left[e^{\widetilde{\varepsilon}_{jt}} \right]$ and $\Upsilon_t = \ln \left[\frac{1}{\sigma_t} + 1 \right]$. The intermediate input elasticity can be rewritten so that we can break it into two parts: a component that varies with inputs and a constant μ, i.e., $\ln D(k_{jt}, l_{jt}, m_{jt}) = \ln D^{\mu}(k_{jt}, l_{jt}, m_{jt}) + \mu$. Then, equation (47) becomes

$$\begin{aligned} s_{jt} &= (\Upsilon_t + \mu) + \ln \widetilde{\mathcal{E}} + \ln D^{\mu}(k_{jt}, l_{jt}, m_{jt}) - \widetilde{\varepsilon}_{jt} \\ &= \varphi_t + \ln \widetilde{\mathcal{E}} + \ln D^{\mu}(k_{jt}, l_{jt}, m_{jt}) - \widetilde{\varepsilon}_{jt}. \end{aligned} \tag{48}$$

As equation (48) makes clear, without observing prices, we can nonparametrically recover the scaled ex-post shock $\widetilde{\varepsilon}_{jt}$ (and hence $\widetilde{\mathcal{E}}$); the output elasticity of intermediate inputs up to a constant $\ln D^{\mu}(k_{jt}, l_{jt}, m_{jt}) = \ln D(k_{jt}, l_{jt}, m_{jt}) - \mu$; and the time-varying markups up to the same constant, $\varphi_t = \Upsilon_t + \mu$, using time dummies for φ_t. Recovering the growth pattern of markups over time is useful as an independent result as it can, for example, be used to check whether market power has increased over time, or to analyze the behavior of market power with respect to the business cycle.

As before, we can correct our estimates for $\widetilde{\mathcal{E}}$ and solve the differential equation that arises from equation (48). However, because we can still only identify the elasticity up to the constant μ, we have to be careful about keeping track of it as we can only calculate $\int D^{\mu}(k_{jt}, l_{jt}, m_{jt})\, dm_{jt} = e^{-\mu} \int D(k_{jt}, l_{jt}, m_{jt})\, dm_{jt}$. It follows that

$$f(k_{jt}, l_{jt}, m_{jt})\, e^{-\mu} + \mathscr{C}(k_{jt}, l_{jt})\, e^{-\mu} = D^{\mu}(k_{jt}, l_{jt}, m_{jt})\, dm_{jt}.$$

From this equation it is immediately apparent that, without further information, we will not be able to separate the integration constant $\mathscr{C}(k_{jt}, l_{jt})$ from the unknown constant μ.

To see how both the constant μ and the constant of integration can be recovered, notice that what we observe in the data is the firm's real revenue, which in logs is given by $r_{jt} = (p_{jt} - \pi_t) + y_{jt}$. Recalling equation (1), and replacing for $p_{jt} - \pi_t$ using (45), the observed log-revenue production function is

$$r_{jt} = \left(\frac{1}{\sigma_t} + 1\right) f(k_{jt}, l_{jt}, m_{jt}) - \frac{1}{\sigma_t} y_t + \chi_{jt} + \left(\frac{1}{\sigma_t} + 1\right) \omega_{jt} + \widetilde{\varepsilon}_{jt}. \tag{49}$$

However, we can write $1 + \frac{1}{\sigma_t} = e^{\varphi_t} e^{-\mu}$. We know φ_t from our analysis above, so only μ is unknown. Replacing back into (49) we get

$$\begin{aligned} r_{jt} &= e^{\varphi_t} e^{-\mu} f\left(k_{jt}, l_{jt}, m_{jt}\right) - \left(e^{\varphi_t} e^{-\mu} - 1\right) y_t \\ &+ e^{\varphi_t} e^{-\mu} \left(\omega_{jt} + \chi_{jt}\right) + \widetilde{\varepsilon}_{jt}. \end{aligned} \tag{50}$$

We then follow a similar strategy as before. As in equation (16), we first form an observable variable

$$\mathcal{R}_{jt} \equiv \ln \frac{\frac{P_{jt} Y_{jt}}{\Pi_t}}{e^{\widetilde{\varepsilon}_{jt}} e^{e^{\varphi_t} \int D^{\mu}(k_{jt}, l_{jt}, m_{jt}) dm_{jt}}},$$

where we now use revenues (the measure of output we observe), include e^{φ_t}, as well as use D^{μ} instead of the (for now) unobservable D. Replacing into (50) we obtain

$$\mathcal{R}_{jt} = -e^{\varphi_t - \mu} \mathscr{C}\left(k_{jt}, l_{jt}\right) - \left(e^{\varphi_t} e^{-\mu} - 1\right) y_t + e^{\varphi_t} e^{-\mu} \left(\omega_{jt} + \chi_{jt}\right).$$

From this equation it is clear that the constant μ will be identified from variation in the observed demand shifter y_t. Without having recovered φ_t from the share regression first, it would not be possible to identify time-varying markups. Note that in equation (49), both σ_t and y_t change with time, and hence y_t cannot be used to identify σ_t unless we restrict $\sigma_t = \sigma$ as in Klette and Griliches (1996) and De Loecker (2011).

Finally, we can only recover a linear combination of productivity and the demand shock, $1 + \frac{1}{\sigma_t} \left(\omega_{jt} + \chi_{jt}\right)$. The reason is clear: since we do not observe prices, we have no way of disentangling whether, after controlling for inputs, a firm has higher revenues because it is more productive (ω_{jt}) or because it can sell at a higher price (χ_{jt}). We can write $\omega_{jt}^{\mu} = 1 + \frac{1}{\sigma_t} \left(\omega_{jt} + \chi_{jt}\right)$ as a function

of the parts that remain to be recovered

$$\omega_{jt}^{\mu} = \mathcal{R}_{jt} + e^{\varphi_t - \mu}\mathscr{C}\left(k_{jt}, l_{jt}\right) + \quad e^{\varphi_t}e^{-\mu} - 1 \quad y_t,$$

and impose the Markovian assumption on this combination:[64] $\omega_{jt}^{\mu} = h \quad \omega_{jt-1}^{\mu} \quad + \eta_{jt}^{\mu}$. We can use similar moment restrictions as before, $E \quad \eta_{jt}^{\mu} | k_{jt}, l_{jt} \quad = 0$, to identify the constant of integration $\mathscr{C}\left(k_{jt}, l_{jt}\right)$ as well as μ (and hence the level of the markups).

[64]Note that in general the sum of two first-order Markov processes is not a first-order Markov process itself. In this case, one would need to replace Assumption 2 with the assumption that the weighted sum of productivity ω_{jt} and the demand shock χ_{jt} is Markovian. See De Loecker (2011) for an example that imposes this assumption.

Online Appendix O6: Additional Results

Table O6.1: Average Input Elasticities of Output--Energy+Services Flexible
(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	0.15	0.13	0.21	0.16	0.37	0.31	0.28	0.27	0.29	0.26	0.22	0.21
	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.06)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
Capital	0.06	0.03	0.05	0.04	0.05	0.02	0.01	0.02	0.04	0.02	0.08	0.05
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
Raw Materials	0.71	0.67	0.69	0.55	0.49	0.50	0.63	0.46	0.58	0.53	0.63	0.53
	(0.02)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.07)	(0.02)	(0.03)	(0.01)	(0.01)	(0.01)
Energy+Services	0.08	0.18	0.11	0.27	0.09	0.18	0.10	0.23	0.11	0.23	0.11	0.24
	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.02)	(0.00)	(0.01)	(0.00)	(0.00)
Sum	1.01	1.01	1.05	1.01	1.00	1.00	1.02	0.98	1.02	1.04	1.04	1.02
	(0.01)	(0.00)	(0.02)	(0.01)	(0.01)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)
Mean(Capital) / Mean(Labor)	0.43	0.26	0.23	0.24	0.12	0.05	0.04	0.07	0.14	0.08	0.37	0.22
	(0.08)	(0.07)	(0.14)	(0.07)	(0.04)	(0.02)	(0.08)	(0.05)	(0.06)	(0.04)	(0.04)	(0.02)
Chile												
Labor	0.18	0.14	0.28	0.22	0.31	0.25	0.29	0.20	0.31	0.30	0.22	0.18
	(0.02)	(0.01)	(0.03)	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Capital	0.06	0.04	0.08	0.05	0.04	0.03	0.06	0.03	0.08	0.06	0.11	0.08
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Raw Materials	0.77	0.72	0.65	0.62	0.65	0.64	0.59	0.65	0.63	0.58	0.67	0.63
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.05)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
Energy+Services	0.07	0.14	0.07	0.16	0.06	0.13	0.11	0.17	0.07	0.15	0.07	0.16
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.00)
Sum	1.08	1.04	1.08	1.05	1.06	1.05	1.05	1.04	1.10	1.09	1.08	1.05
	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Mean(Capital) / Mean(Labor)	0.36	0.27	0.28	0.21	0.13	0.13	0.20	0.12	0.26	0.20	0.49	0.42
	(0.04)	(0.03)	(0.05)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification in which energy+services is flexible and raw materials is not flexible. In the first column the results are obtained via our approach (labeled GNR) using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ). The numbers in the second column are estimated using a complete polynomial series of degree 2 with OLS.

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, raw materials, and energy+services elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

Table O6.2: Heterogeneity in Productivity--Energy+Services Flexible
(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	1.20	1.18	1.25	1.21	1.24	1.19	1.30	1.25	1.28	1.21	1.33	1.24
	(0.02)	(0.01)	(0.03)	(0.01)	(0.03)	(0.01)	(0.06)	(0.02)	(0.02)	(0.01)	(0.01)	(0.00)
90/10 ratio	1.50	1.45	1.62	1.51	1.60	1.49	1.75	1.57	1.68	1.52	1.83	1.60
	(0.05)	(0.03)	(0.10)	(0.03)	(0.07)	(0.02)	(0.16)	(0.06)	(0.06)	(0.02)	(0.03)	(0.01)
95/5 ratio	1.87	1.80	2.09	1.85	2.00	1.80	2.26	2.00	2.04	1.79	2.43	2.00
	(0.09)	(0.07)	(0.22)	(0.07)	(0.11)	(0.04)	(0.24)	(0.13)	(0.11)	(0.04)	(0.08)	(0.02)
Exporter	0.14	0.11	-0.04	-0.02	0.02	0.01	0.14	0.07	0.08	0.01	0.01	0.00
	(0.04)	(0.04)	(0.06)	(0.01)	(0.03)	(0.02)	(0.12)	(0.09)	(0.02)	(0.01)	(0.03)	(0.01)
Importer	0.00	-0.03	-0.03	-0.01	-0.03	0.00	-0.04	-0.06	0.10	0.04	-0.02	0.01
	(0.02)	(0.01)	(0.06)	(0.01)	(0.03)	(0.01)	(0.05)	(0.02)	(0.02)	(0.01)	(0.05)	(0.01)
Advertiser	-0.12	-0.10	-0.13	-0.05	-0.10	-0.07	-0.07	-0.04	0.05	-0.02	-0.16	-0.05
	(0.03)	(0.02)	(0.11)	(0.02)	(0.04)	(0.02)	(0.06)	(0.03)	(0.02)	(0.01)	(0.05)	(0.01)
Wages > Median	0.06	0.04	0.13	0.08	0.14	0.12	0.09	0.08	0.19	0.11	0.10	0.10
	(0.02)	(0.02)	(0.05)	(0.01)	(0.02)	(0.01)	(0.05)	(0.03)	(0.02)	(0.01)	(0.04)	(0.01)
Chile												
75/25 ratio	1.31	1.29	1.42	1.38	1.41	1.36	1.44	1.39	1.48	1.45	1.49	1.44
	(0.01)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)	(0.02)	(0.01)	(0.01)	(0.00)
90/10 ratio	1.76	1.71	2.04	1.94	2.01	1.89	2.13	2.03	2.23	2.12	2.26	2.12
	(0.03)	(0.01)	(0.05)	(0.03)	(0.04)	(0.02)	(0.12)	(0.04)	(0.05)	(0.04)	(0.02)	(0.01)
95/5 ratio	2.22	2.12	2.69	2.50	2.65	2.42	2.94	2.76	2.94	2.79	3.08	2.83
	(0.05)	(0.03)	(0.12)	(0.05)	(0.07)	(0.04)	(0.21)	(0.08)	(0.08)	(0.07)	(0.04)	(0.02)
Exporter	-0.01	-0.01	-0.06	-0.01	0.01	0.01	0.01	-0.03	-0.05	0.00	-0.03	-0.01
	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.05)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)
Importer	0.02	0.03	0.04	0.03	0.07	0.06	0.09	0.07	0.05	0.06	0.07	0.09
	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.04)	(0.03)	(0.02)	(0.01)	(0.01)	(0.01)
Advertiser	-0.01	-0.01	-0.01	0.00	0.01	0.01	0.01	0.01	0.00	0.01	0.02	0.03
	(0.01)	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Wages > Median	0.12	0.10	0.15	0.13	0.19	0.15	0.17	0.13	0.16	0.15	0.25	0.22
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.03)	(0.01)	(0.03)	(0.02)	(0.01)	(0.01)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification in which energy+services is flexible and raw materials is not flexible. In the first column the results are obtained via our approach (labeled GNR) using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ϕ). The numbers in the second column are estimated using a complete polynomial series of degree 2 with OLS.

c. 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) based on 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, 1% less productive than a firm that does not advertise.

Table O6.3: Average Input Elasticities of Output--Raw Materials Flexible
(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	0.13	0.13	0.21	0.16	0.33	0.31	0.28	0.27	0.30	0.26	0.24	0.21
	(0.02)	(0.01)	(0.04)	(0.02)	(0.03)	(0.01)	(0.06)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)
Capital	0.05	0.03	0.07	0.04	0.02	0.02	0.01	0.02	0.05	0.02	0.06	0.05
	(0.01)	(0.01)	(0.03)	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)	(0.02)	(0.01)	(0.02)	(0.00)
Raw Materials	0.60	0.67	0.44	0.55	0.41	0.50	0.42	0.46	0.42	0.53	0.43	0.53
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)
Energy+Services	0.22	0.18	0.30	0.27	0.23	0.18	0.26	0.23	0.28	0.23	0.29	0.24
	(0.02)	(0.01)	(0.05)	(0.02)	(0.03)	(0.01)	(0.07)	(0.02)	(0.04)	(0.01)	(0.02)	(0.00)
Sum	1.00	1.01	1.01	1.01	1.00	1.00	0.97	0.98	1.05	1.04	1.02	1.02
	(0.01)	(0.00)	(0.04)	(0.01)	(0.03)	(0.01)	(0.05)	(0.02)	(0.04)	(0.01)	(0.01)	(0.00)
Mean(Capital) / Mean(Labor)	0.39	0.26	0.33	0.24	0.07	0.05	0.04	0.07	0.16	0.08	0.26	0.22
	(0.12)	(0.07)	(0.26)	(0.07)	(0.06)	(0.02)	(0.15)	(0.05)	(0.07)	(0.04)	(0.09)	(0.02)
Chile												
Labor	0.19	0.14	0.34	0.22	0.35	0.25	0.38	0.20	0.42	0.30	0.26	0.18
	(0.02)	(0.01)	(0.03)	(0.02)	(0.04)	(0.02)	(0.04)	(0.02)	(0.05)	(0.02)	(0.02)	(0.01)
Capital	0.06	0.04	0.07	0.05	0.06	0.03	0.06	0.03	0.12	0.06	0.11	0.08
	(0.01)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.00)
Raw Materials	0.59	0.72	0.47	0.62	0.49	0.64	0.46	0.65	0.43	0.58	0.47	0.63
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)
Energy+Services	0.21	0.14	0.18	0.16	0.15	0.13	0.16	0.17	0.18	0.15	0.21	0.16
	(0.02)	(0.00)	(0.03)	(0.01)	(0.04)	(0.01)	(0.03)	(0.01)	(0.06)	(0.01)	(0.02)	(0.00)
Sum	1.05	1.04	1.06	1.05	1.05	1.05	1.07	1.04	1.14	1.09	1.05	1.05
	(0.02)	(0.00)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.00)
Mean(Capital) / Mean(Labor)	0.32	0.27	0.20	0.21	0.16	0.13	0.16	0.12	0.28	0.20	0.41	0.42
	(0.04)	(0.03)	(0.06)	(0.04)	(0.04)	(0.04)	(0.06)	(0.04)	(0.07)	(0.04)	(0.04)	(0.03)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification in which raw materials is flexible and energy+services is not flexible. In the first column the results are obtained via our approach (labeled GNR) using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ϕ). The numbers in the second column are estimated using a complete polynomial series of degree 2 with OLS.

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, raw materials, and energy+services elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

Table O6.4: Heterogeneity in Productivity--Raw Materials Flexible
(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	1.24	1.18	1.27	1.21	1.25	1.19	1.24	1.25	1.24	1.21	1.31	1.24
	(0.02)	(0.01)	(0.07)	(0.01)	(0.03)	(0.01)	(0.07)	(0.02)	(0.07)	(0.01)	(0.02)	(0.00)
90/10 ratio	1.58	1.45	1.64	1.51	1.59	1.49	1.62	1.57	1.57	1.52	1.72	1.60
	(0.03)	(0.03)	(0.16)	(0.03)	(0.08)	(0.02)	(0.19)	(0.06)	(0.19)	(0.02)	(0.05)	(0.01)
95/5 ratio	1.92	1.80	2.10	1.85	1.90	1.80	2.01	2.00	1.89	1.79	2.13	2.00
	(0.06)	(0.07)	(0.24)	(0.07)	(0.13)	(0.04)	(0.41)	(0.13)	(0.30)	(0.04)	(0.07)	(0.02)
Exporter	0.09	0.11	-0.01	-0.02	0.04	0.01	0.05	0.07	0.01	0.01	0.04	0.00
	(0.04)	(0.04)	(0.09)	(0.01)	(0.07)	(0.02)	(0.18)	(0.09)	(0.11)	(0.01)	(0.01)	(0.01)
Importer	0.02	-0.03	0.03	-0.01	0.08	0.00	-0.03	-0.06	0.05	0.04	0.07	0.01
	(0.02)	(0.01)	(0.09)	(0.01)	(0.10)	(0.01)	(0.15)	(0.02)	(0.08)	(0.01)	(0.01)	(0.01)
Advertiser	-0.06	-0.10	0.01	-0.05	-0.01	-0.07	-0.02	-0.04	-0.02	-0.02	-0.03	-0.05
	(0.02)	(0.02)	(0.05)	(0.02)	(0.04)	(0.02)	(0.05)	(0.03)	(0.05)	(0.01)	(0.01)	(0.01)
Wages > Median	0.04	0.04	0.11	0.08	0.14	0.12	0.09	0.08	0.11	0.11	0.13	0.10
	(0.02)	(0.02)	(0.07)	(0.01)	(0.03)	(0.01)	(0.06)	(0.03)	(0.07)	(0.01)	(0.01)	(0.01)
Chile												
75/25 ratio	1.34	1.29	1.44	1.38	1.40	1.36	1.50	1.39	1.52	1.45	1.51	1.44
	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)	(0.00)
90/10 ratio	1.82	1.71	2.13	1.94	2.05	1.89	2.31	2.03	2.26	2.12	2.32	2.12
	(0.07)	(0.01)	(0.06)	(0.03)	(0.06)	(0.02)	(0.06)	(0.04)	(0.13)	(0.04)	(0.02)	(0.01)
95/5 ratio	2.32	2.12	2.80	2.50	2.70	2.42	3.09	2.76	2.98	2.79	3.19	2.83
	(0.12)	(0.03)	(0.10)	(0.05)	(0.10)	(0.04)	(0.11)	(0.08)	(0.25)	(0.07)	(0.04)	(0.02)
Exporter	-0.02	-0.01	0.01	-0.01	0.05	0.01	-0.01	-0.03	-0.03	0.00	0.01	-0.01
	(0.06)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)	(0.01)	(0.01)
Importer	0.06	0.03	0.06	0.03	0.09	0.06	0.12	0.07	0.06	0.06	0.13	0.09
	(0.06)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.04)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)
Advertiser	0.00	-0.01	0.02	0.00	0.02	0.01	0.03	0.01	-0.02	0.01	0.04	0.03
	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.01)	(0.01)
Wages > Median	0.13	0.10	0.15	0.13	0.18	0.15	0.19	0.13	0.17	0.15	0.25	0.22
	(0.04)	(0.01)	(0.03)	(0.02)	(0.03)	(0.01)	(0.02)	(0.01)	(0.04)	(0.02)	(0.01)	(0.01)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification in which raw materials is flexible and energy+services is not flexible. In the first column the results are obtained via our approach (labeled GNR) using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and $\mathcal{E}$). The numbers in the second column are estimated using a complete polynomial series of degree 2 with OLS.

c. 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) based on 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, 0% less productive than a firm that does not advertise.

Table O6.5: Average Input Elasticities of Output--Fixed Effects

(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
Labor	0.18	0.26	0.39	0.46	0.29	0.33
	(0.05)	(0.07)	(0.04)	(0.12)	(0.09)	(0.02)
Capital	0.09	0.04	0.06	0.25	0.09	0.07
	(0.07)	(0.06)	(0.04)	(0.16)	(0.11)	(0.02)
Intermediates	0.67	0.54	0.52	0.51	0.53	0.54
	(0.01)	(0.01)	(0.01)	(0.02)	(0.01)	(0.00)
Sum	0.95	0.84	0.96	1.22	0.90	0.95
	(0.12)	(0.11)	(0.07)	(0.26)	(0.18)	(0.04)
Mean(Capital) / Mean(Labor)	0.52	0.16	0.15	0.55	0.30	0.21
	(1.42)	(0.66)	(0.09)	(0.34)	(0.35)	(0.07)
Chile						
Labor	0.20	0.33	0.50	0.37	0.60	0.30
	(0.03)	(0.07)	(0.05)	(0.03)	(0.15)	(0.02)
Capital	0.02	0.08	0.17	0.10	0.32	0.15
	(0.06)	(0.09)	(0.07)	(0.06)	(0.15)	(0.05)
Intermediates	0.67	0.54	0.56	0.59	0.50	0.55
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Sum	0.89	0.95	1.24	1.05	1.42	1.01
	(0.08)	(0.13)	(0.11)	(0.07)	(0.29)	(0.07)
Mean(Capital) / Mean(Labor)	0.09	0.23	0.35	0.27	0.53	0.50
	(0.25)	(0.23)	(0.11)	(0.14)	(0.35)	(0.15)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification with fixed effects and are estimated using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ) of our approach (labeled GNR).

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, and intermediate input elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

Table O6.6: Heterogeneity in Productivity--Fixed Effects
(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
75/25 ratio	1.36	1.67	1.30	1.58	1.52	1.52
	(0.32)	(0.40)	(0.06)	(0.46)	(0.32)	(0.08)
90/10 ratio	1.82	2.82	1.70	2.71	2.20	2.21
	(1.25)	(1.71)	(0.18)	(2.82)	(1.04)	(0.25)
95/5 ratio	2.30	4.14	2.09	4.04	2.78	2.84
	(2.66)	(4.01)	(0.35)	(13.90)	(1.75)	(0.41)
Exporter	0.26	0.25	0.10	-0.04	0.41	0.22
	(0.35)	(0.95)	(0.19)	(2.64)	(0.50)	(0.11)
Importer	0.13	0.35	0.19	-0.08	0.32	0.27
	(0.29)	(0.76)	(0.25)	(2.25)	(0.37)	(0.10)
Advertiser	0.01	0.32	0.07	-0.17	0.19	0.14
	(0.09)	(0.30)	(0.08)	(0.41)	(0.24)	(0.04)
Wages > Median	0.17	0.45	0.20	0.02	0.40	0.37
	(0.26)	(0.53)	(0.06)	(0.51)	(0.33)	(0.09)
Chile						
75/25 ratio	1.57	1.60	1.52	1.52	2.06	1.57
	(0.15)	(0.17)	(0.12)	(0.13)	(0.36)	(0.15)
90/10 ratio	2.41	2.55	2.40	2.34	4.48	2.45
	(0.40)	(0.59)	(0.45)	(0.48)	(1.27)	(0.45)
95/5 ratio	3.14	3.38	3.38	3.20	7.30	3.41
	(0.61)	(1.20)	(0.98)	(0.99)	(2.55)	(0.77)
Exporter	0.34	0.07	-0.07	0.07	-0.42	0.14
	(0.23)	(0.21)	(0.12)	(0.42)	(0.38)	(0.24)
Importer	0.51	0.17	-0.02	0.22	-0.25	0.25
	(0.26)	(0.18)	(0.11)	(0.31)	(0.39)	(0.21)
Advertiser	0.22	0.13	-0.06	0.04	-0.20	0.12
	(0.11)	(0.14)	(0.09)	(0.07)	(0.23)	(0.10)
Wages > Median	0.50	0.28	0.10	0.23	-0.12	0.39
	(0.20)	(0.17)	(0.08)	(0.15)	(0.44)	(0.22)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification with fixed effects and are estimated using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ) of our approach (labeled GNR).

c. 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) based on 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 a firm that advertises is, on average, 22% more productive than a firm that does not advertise.

Table O6.7: Average Input Elasticities of Output--Extra Unobservable
(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
Labor	0.18	0.32	0.39	0.45	0.40	0.36
	(0.04)	(0.04)	(0.03)	(0.07)	(0.03)	(0.01)
Capital	0.13	0.18	0.08	-0.01	0.11	0.15
	(0.03)	(0.02)	(0.02)	(0.04)	(0.02)	(0.01)
Intermediates	0.67	0.54	0.52	0.51	0.53	0.54
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Sum	0.98	1.03	0.99	0.95	1.03	1.05
	(0.02)	(0.03)	(0.02)	(0.08)	(0.02)	(0.01)
Mean(Capital) / Mean(Labor)	0.72	0.55	0.21	-0.02	0.27	0.41
	(0.65)	(0.13)	(0.09)	(0.09)	(0.05)	(0.03)
Chile						
Labor	0.24	0.44	0.45	0.37	0.52	0.36
	(0.01)	(0.03)	(0.02)	(0.03)	(0.03)	(0.01)
Capital	0.12	0.11	0.07	0.09	0.14	0.17
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Intermediates	0.66	0.54	0.55	0.59	0.50	0.55
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Sum	1.02	1.09	1.08	1.04	1.16	1.08
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Mean(Capital) / Mean(Labor)	0.50	0.26	0.15	0.25	0.28	0.48
	(0.05)	(0.05)	(0.04)	(0.05)	(0.04)	(0.02)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification allowing for an extra unobservable in the share equation. The estimates are obtained using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and $\mathcal{E}$) of our approach (labeled GNR).

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, and intermediate input elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

Table O6.8: Heterogeneity in Productivity--Extra Unobservable
(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
75/25 ratio	1.35	1.35	1.29	1.35	1.32	1.37
	(0.04)	(0.03)	(0.02)	(0.08)	(0.03)	(0.01)
90/10 ratio	1.82	1.83	1.68	1.94	1.76	1.88
	(0.13)	(0.08)	(0.06)	(0.30)	(0.05)	(0.02)
95/5 ratio	2.36	2.34	2.03	2.57	2.18	2.37
	(0.26)	(0.17)	(0.11)	(0.70)	(0.09)	(0.03)
Exporter	0.15	0.03	0.04	0.32	0.11	0.06
	(0.07)	(0.04)	(0.04)	(0.25)	(0.04)	(0.01)
Importer	0.03	0.05	0.11	0.14	0.12	0.11
	(0.04)	(0.04)	(0.04)	(0.15)	(0.03)	(0.01)
Advertiser	-0.03	0.05	0.03	0.07	0.07	0.02
	(0.03)	(0.04)	(0.03)	(0.11)	(0.03)	(0.01)
Wages > Median	0.09	0.18	0.18	0.21	0.23	0.19
	(0.04)	(0.04)	(0.02)	(0.10)	(0.03)	(0.01)
Chile						
75/25 ratio	1.37	1.49	1.43	1.51	1.54	1.55
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
90/10 ratio	1.92	2.18	2.12	2.35	2.35	2.40
	(0.03)	(0.09)	(0.04)	(0.05)	(0.06)	(0.02)
95/5 ratio	2.51	2.93	2.77	3.15	3.12	3.33
	(0.06)	(0.18)	(0.08)	(0.11)	(0.12)	(0.04)
Exporter	0.00	0.03	0.09	0.00	-0.02	0.03
	(0.04)	(0.05)	(0.03)	(0.04)	(0.03)	(0.01)
Importer	0.12	0.10	0.13	0.15	0.09	0.15
	(0.04)	(0.04)	(0.02)	(0.04)	(0.03)	(0.01)
Advertiser	0.04	0.04	0.06	0.04	0.01	0.06
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Wages > Median	0.20	0.19	0.22	0.22	0.20	0.30
	(0.03)	(0.04)	(0.02)	(0.03)	(0.03)	(0.01)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification allowing for an extra unobservable in the share equation. The estimates are obtained using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and $\mathcal{E}$) of our approach (labeled GNR).

c. 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) based on 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 a firm that advertises is, on average, 4% more productive than a firm that does not advertise.

Table O6.9: Average Input Elasticities of Output--Ex-post Shock Robustness
(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
Labor	0.21	0.32	0.42	0.45	0.43	0.35
	(0.02)	(0.03)	(0.02)	(0.04)	(0.02)	(0.01)
Capital	0.12	0.15	0.05	0.04	0.10	0.14
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Intermediates	0.67	0.54	0.52	0.51	0.53	0.54
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Sum	1.01	1.01	0.99	0.99	1.06	1.04
	(0.01)	(0.02)	(0.01)	(0.04)	(0.01)	(0.01)
Mean(Capital) / Mean(Labor)	0.59	0.49	0.12	0.08	0.23	0.40
	(0.09)	(0.09)	(0.04)	(0.06)	(0.04)	(0.02)
Chile						
Labor	0.28	0.45	0.45	0.40	0.52	0.38
	(0.01)	(0.03)	(0.02)	(0.02)	(0.02)	(0.01)
Capital	0.11	0.11	0.06	0.07	0.13	0.16
	(0.01)	(0.02)	(0.01)	(0.01)	(0.01)	(0.00)
Intermediates	0.67	0.54	0.56	0.59	0.50	0.55
	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Sum	1.05	1.10	1.08	1.06	1.15	1.09
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Mean(Capital) / Mean(Labor)	0.39	0.24	0.14	0.18	0.25	0.43
	(0.02)	(0.05)	(0.03)	(0.03)	(0.03)	(0.02)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification allowing for the expectation of e^{ε} to depend on capital, labor, lagged intermediate inputs, and time. The estimates are obtained using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ) of our approach (labeled GNR).

c. Since the input elasticities are heterogeneous across firms, we report the average input elasticities within each given industry.

d. The row titled "Sum" reports the sum of the average labor, capital, and intermediate input elasticities, and the row titled "Mean(Capital)/Mean(Labor)" reports the ratio of the average capital elasticity to the average labor elasticity.

Table O6.10: Heterogeneity in Productivity--Ex-post Shock Robustness
(Structural Estimates)

	Food Products (311)	Textiles (321)	Apparel (322)	Wood Products (331)	Fabricated Metals (381)	All
	GNR	GNR	GNR	GNR	GNR	GNR
Colombia						
75/25 ratio	1.35	1.35	1.29	1.30	1.31	1.37
	(0.03)	(0.03)	(0.01)	(0.03)	(0.02)	(0.01)
90/10 ratio	1.82	1.83	1.66	1.80	1.75	1.87
	(0.07)	(0.07)	(0.03)	(0.12)	(0.04)	(0.02)
95/5 ratio	2.29	2.39	2.03	2.25	2.15	2.36
	(0.13)	(0.15)	(0.05)	(0.24)	(0.06)	(0.03)
Exporter	0.14	0.02	0.05	0.15	0.08	0.06
	(0.05)	(0.04)	(0.03)	(0.17)	(0.03)	(0.01)
Importer	0.03	0.05	0.12	0.04	0.10	0.11
	(0.03)	(0.04)	(0.03)	(0.11)	(0.02)	(0.01)
Advertiser	-0.03	0.08	0.05	0.04	0.05	0.03
	(0.02)	(0.03)	(0.02)	(0.05)	(0.02)	(0.01)
Wages > Median	0.09	0.18	0.18	0.15	0.22	0.20
	(0.03)	(0.03)	(0.01)	(0.05)	(0.02)	(0.01)
Chile						
75/25 ratio	1.38	1.48	1.43	1.50	1.53	1.55
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
90/10 ratio	1.91	2.16	2.11	2.32	2.32	2.39
	(0.02)	(0.04)	(0.04)	(0.05)	(0.05)	(0.02)
95/5 ratio	2.48	2.92	2.77	3.11	3.12	3.31
	(0.05)	(0.07)	(0.08)	(0.09)	(0.10)	(0.04)
Exporter	0.02	0.02	0.09	0.00	-0.01	0.03
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.01)
Importer	0.15	0.10	0.14	0.15	0.10	0.15
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.01)
Advertiser	0.04	0.04	0.06	0.03	0.01	0.06
	(0.01)	(0.02)	(0.02)	(0.01)	(0.02)	(0.01)
Wages > Median	0.22	0.19	0.23	0.21	0.22	0.30
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)	(0.01)

Notes:

a. Standard errors are estimated using the bootstrap with 200 replications and are reported in parentheses below the point estimates.

b. For each industry, the numbers are based on a gross output specification allowing for the expectation of e^{ε} to depend on capital, labor, lagged intermediate inputs, and time. The estimates are obtained using a complete polynomial series of degree 2 for each of the two nonparametric functions (D and ℓ) of our approach (labeled GNR).

c. 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) based on 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 a firm that advertises is, on average, 4% more productive than a firm that does not advertise.

References Online Appendix

Cunha, Flavio, James J. Heckman, and Susanne M. Schennach, "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 05 2010, *78* (3), 883–931.

Davidson, Russell and James G MacKinnon, *Econometric Theory and Methods*, Vol. 5, New York: Oxford University Press, 2004.

Doraszelski, Ulrich and Jordi Jaumandreu, "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies*, 2013, *80* (4), 1338–1383.

_ **and** _ , "Measuring the Bias of Technological Change," 2015. Working Paper.

Hu, Yingyao and Susanne M Schennach, "Instrumental variable treatment of nonclassical measurement error models," *Econometrica*, 2008, *76* (1), 195–216.

Hurwicz, Leonid, "Generalization of the Concept of Identification," in "Statistical Inference in Dynamic Economc Models, Cowles Commission Monograph 10," New York: John Wiley, 1950.

Judd, Kenneth L, *Numerical Methods in Economics*, Cambridge: MIT Press, 1998.

_ **, Paul Schrimpf, and Michio Suzuki**, "Identification and Estimation of Production Function with Unobserved Heterogeneity," 2015. Working Paper.

Klette, Tor Jacob and Zvi Griliches, "The Inconsistency of Common Scale Estimators When Output Prices are Unobserved and Endogenous," *Journal of Applied Econometrics*, July 1996, *11* (4), 343–361.

Loecker, Jan De, "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity," *Econometrica*, 2011, *79* (5), 1407–1451.

Matzkin, Rosa L., "Nonparametric Identification," in James J. Heckman and Edward E. Leamer, eds., *Handbook of Econometrics*, Vol. 6, Amsterdam: Elsevier, 2007, pp. 5307–5368.

_ , "Nonparametric Identification in Structural Economic Models," *Annual Review of Economics*, 2013, *5* (1), 457–486.

_ **and James L. Powell**, "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 2003, *71* (5), 1565–1578.

Varian, Hal R, *Microeconomic Analysis*, New York: WW Norton, 1992.