

NBER WORKING PAPER SERIES

ESTIMATING PRODUCTION FUNCTIONS IN DIFFERENTIATED-PRODUCT
INDUSTRIES WITH QUANTITY INFORMATION AND EXTERNAL INSTRUMENTS

Nicolás de Roux
Marcela Eslava
Santiago Franco
Eric Verhoogen

Working Paper 28323
<http://www.nber.org/papers/w28323>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
January 2021, Revised July 2023

We thank David Atkin, Mons Chan, Rafael Dix-Carneiro, Ulrich Doraszelski, Dave Donaldson, Garth Frazer, Álvaro García-Marin, Arshia Hashemi, Hiro Kasahara, Amit Khandelwal, Pete Klenow, Sebastian Kripfganz, Jeanne LaFortune, Suresh Naidu, Scott Orr, Amil Petrin, David Rivers, José Tessada, James Traina, Jim Tybout, Johannes van Biesebroeck, Reed Walker, Frederic Warzynski, Daniel Xu, and participants in several seminars for helpful comments, and Nathalie González and Cristhian Acosta for excellent research assistance. We are grateful to DANE staff for facilitating data access and for advice on the proper use of the data. The views expressed in this document are not those of DANE or DANE staff. The analysis using confidential data was carried out in accordance with Colombian confidentiality laws. All errors are our own. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2021 by Nicolás de Roux, Marcela Eslava, Santiago Franco, and Eric Verhoogen. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating Production Functions in Differentiated-Product Industries with Quantity Information and External Instruments

Nicolás de Roux, Marcela Eslava, Santiago Franco, and Eric Verhoogen

NBER Working Paper No. 28323

January 2021, Revised July 2023

JEL No. D24,L1,L65,O14

ABSTRACT

This paper highlights shortcomings of standard methods of production-function estimation when quality or variety vary at the firm level and develops a new approach that can be applied in such contexts. We take advantage of input and output quantity data from Colombian producers of rubber and plastic products. Using constant-elasticity-of-substitution aggregators of outputs and material inputs at the firm level, we derive a simple expression showing how quality and variety choices may bias standard estimators. Using real exchange rates and variation in the “bite” of the national minimum wage, we construct external instruments for materials and labor choices to supplement standard internal instruments. We implement a two-step instrumental-variables method, estimating a difference equation to recover the materials and labor coefficients and then a levels equation to recover the capital coefficient. A simple Monte-Carlo simulation illustrates the advantages of our method in a setting with firm-level input-quality differences.

Nicolás de Roux
Department of Economics
Universidad de los Andes
Calle 19A No 1-37 Este, W 808
Bogotá, Colombia
nicolas.de.roux@uniandes.edu.co

Santiago Franco
University of Chicago
Kenneth C. Griffin Department of Economics
1126 E. 59th St.
Chicago, IL 60637
sfranco@uchicago.edu

Marcela Eslava
Department of Economics
Universidad de los Andes
Calle 19A # 1-37 Este. W 810
Bogotá, Colombia
meslava@uniandes.edu.co

Eric Verhoogen
Department of Economics
School of International & Public Affairs
Columbia University
420 W. 118th Street
New York, NY 10027
and IZA
and also NBER
eric.verhoogen@columbia.edu

A data appendix is available at <http://www.nber.org/data-appendix/w28323>

1 Introduction

A central challenge in estimating production functions is to estimate the elasticities of real output with respect to real inputs, unconfounded by differences in prices across firms. Estimates of these elasticities are key to constructing standard measures of total factor productivity (TFP), the most commonly used metric of firm performance. They are also important for estimating markups in the influential method of Hall (1988) and De Loecker and Warzynski (2012). As recently emphasized by Bond et al. (2021), that method requires an elasticity of real output, not of sales or value-added, in order to generate informative estimates of markups.

Two difficulties in estimating such elasticities have received particular attention. First, prices and physical quantities are usually not observed at the firm-product level. The most common approach is to regress sales (or value-added), deflated by a sector-level price index, on material expenditures and other inputs, similarly deflated. It has long been recognized that the resulting estimates may reflect idiosyncratic variation in market power at the firm level (Klette and Griliches, 1996; Foster et al., 2008, 2016; De Loecker and Goldberg, 2014). Second, firms may choose variable inputs after observing shocks to their own productivity in a given period, generating a positive correlation between unobserved productivity shocks and the input choices — the familiar “transmission bias” problem, first noted by Marschak and Andrews (1944).¹

Information on prices and quantities at the firm-product level, while still uncommon, is increasingly available and has enabled progress on the first issue. For instance, Foster et al. (2008) credibly estimate output elasticities using physical quantities in homogeneous industries in the US Census of Manufactures. But as suggested by Katayama et al. (2009), Grieco and McDevitt (2016), Atkin et al. (2019), and others, using physical quantities may be misleading in differentiated-product industries where the quality and variety of outputs and inputs vary across firms and over time. If product quality and variety are valued by consumers, they should be incorporated in our notion of real output; similarly, if input quality and variety matter for real output, they should be incorporated in real inputs. But once one accepts these propositions, estimates using only physical units may be subject to what we call *quality* and *variety biases*. For example, if producing higher-quality goods requires more labor hours per physical unit of output, the labor coefficient from an OLS regression of physical output on hours (and other inputs) will be biased downwards — it will understate the contribution of labor to real (i.e. quality-adjusted) output. Similar biases can result from changes in output or input variety or input quality, if firms’ choices of physical units of inputs respond.

In this paper, we develop a new approach to estimating output elasticities that takes advantage of quantity information, that is arguably not subject to quality or variety biases, and that also addresses the transmission-bias problem. The method can be applied in horizontally and vertically

¹For reviews, see Griliches and Mairesse (1998), Bartelsman and Doms (2000), Akerberg et al. (2007), De Loecker and Goldberg (2014), Section 2.2.1 of Verhoogen (forthcoming), and De Loecker and Syverson (2021).

differentiated industries with multi-product firms, requires relatively weak assumptions on market structure, and can be used in contexts with time-invariant firm-level heterogeneity in productivity, quality and/or variety. We implement it in data from the Colombian manufacturing survey, which contains information on prices and quantities of both inputs and outputs, focusing (for reasons discussed below) on producers of rubber and plastic products.

We seek to make three main contributions. The first is to highlight conceptually how estimates of output elasticities based on physical quantities may be misleading when quality and variety vary across firms and over time. As in almost all similar datasets, the mapping between specific inputs and specific outputs within the firm is unobserved in the Colombian data. Our approach is to aggregate from the firm-product to the firm level, for both outputs and material inputs. It is not possible to do this aggregation in a theory-free way; any aggregation embeds assumptions, implicit or explicit, about consumer and firm behavior. Here we assume that outputs and inputs, respectively, have constant elasticities of substitution (CES) within firms. We place minimal constraints on substitution elasticities across firms. Following common practice, we assume that (firm-level aggregate) materials, labor, and capital combine in Cobb-Douglas fashion. Although restrictive, the within-firm CES structure is convenient in that it allows us to express the change in each firm-level aggregate as the sum of an observable quantity index and unobservable terms capturing quality and variety. This in turn makes transparent how differences in quality and variety may bias standard estimates. Empirically, our estimates are robust to using other common aggregators at the firm level.

Our second contribution is to address transmission bias by introducing “external” instruments capturing exogenous variation in input prices at the firm level. The idea that external instruments in general, and input prices in particular, would be an attractive solution to the transmission-bias problem has been “in the air” for many years, at least since the landmark review by Griliches and Mairesse (1998).² Several recent papers have acknowledged that factor prices would be natural instruments, but have argued that it would be difficult to find truly exogenous variation at the firm level.³ In the absence of credible external instruments, two approaches have dominated the recent literature. One has been to construct a proxy for unobserved productivity by inverting either an investment-demand or a materials-demand equation, which requires a monotonic relationship between the productivity term (assumed to be scalar) and investment or materials, conditional on other observables (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Wooldridge, 2009; De Loecker,

²Griliches and Mairesse (1998) write that the “future” of production-function estimation lies in “finding circumstances and data that will enable credible identification... The challenge is to find (instrumental) variables that have genuine information about factors which affect firms differentially as they choose their input levels” (p. 198). They describe using “factor prices ... as instrumental variables to identify the parameters of interest” as an “obvious” solution.

³For instance, Akerberg et al. (2015, p. 2418, fn 3) write: “if one observed exogenous, across-firm-variation in all input prices, estimating the production function using input price based IV methods might be preferred to OP/LP [Olley-Pakes/Levinsohn-Petrin] related methodology (due to fewer auxiliary assumptions).” But they also note that “the premise of most of this [proxy-variable] literature is that such variables are either not available or not believed to be exogenous.” See also Akerberg et al. (2007, p. 4208) and Gandhi et al. (2020, Sec. VI.A).

2011; Doraszelski and Jaumandreu, 2013, 2018; Akerberg et al., 2015; Eslava et al., forthcoming). Another approach has been to use “internal” instruments using lagged values of inputs (Chamberlain, 1982; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998, 2000).⁴ In this panel-data approach, the most successful strategy has been the “System GMM” estimator of Arellano and Bover (1995) and Blundell and Bond (1998, 2000), which supplements an equation in first differences, using lagged levels as instruments, with an equation in levels, using lagged differences as instruments.

Our aggregation strategy requires firm-specific normalizations, which we absorb with firm fixed effects. As a consequence, standard proxy-variable approaches are not attractive options for us, because the firm effects would violate the required monotonicity assumption (Akerberg et al., 2015). We instead build on the panel-data approach, which more easily accommodates the firm effects. But rather than include further and further lags as instruments, which is commonly done but may raise weak-instrument concerns, we include two external instruments capturing arguably exogenous variation in input prices, to supplement a parsimonious set of lags. To construct the materials-price instrument, we first use exchange rates to predict import-price changes at the product level, omitting one firm at a time in making the predictions. We then use the lagged product composition of a firm’s imports, in conjunction with the “leave one out” predicted import-price changes, to construct a firm-specific predicted import price index. To construct the wage instrument, we interact changes in the national minimum wage, which saw large increases in real terms over our study period, with an indicator for how binding the change was likely to be on particular firms (the “bite” of the minimum wage), defined as the lagged ratio of the minimum wage to the average wage for permanent employees in the firm. We will see that the external instruments are helpful in alleviating (if not completely removing) weak-instrument concerns.⁵

Our third contribution is a novel approach to estimating the coefficient on capital. It is well known that methods with transformations to remove firm effects — either first-differencing or deviating from firm means — tend to yield implausibly low estimates of the capital coefficient.⁶ The most common explanation is that transformations to remove the firm effect exacerbate attenuation bias due to measurement error (Griliches and Mairesse, 1998). Others have found that this problem persists when instrumenting the change in capital with lagged levels (see e.g. Ornaghi (2006)). A possible reason is that levels of capital stock are particularly poorly measured, even relative to (also noisy) investment (Tybout, 1992; Collard-Wexler and De Loecker, 2020). To improve the estimation of the capital coefficient, we follow the System GMM approach in combining a difference equation, using lagged

⁴The two dominant approaches are themselves related, since the proxy-variable methods often use lagged levels as instruments, as discussed in Akerberg et al. (2015, Sec. 4.3.3).

⁵An additional motivation for including the external instruments is provided by a recent paper by Andrews et al. (2022): under certain conditions, estimates using external instruments maintain a causal interpretation even in the presence of model mis-specification.

⁶See e.g. Griliches and Mairesse (1998), Akerberg et al. (2007), and Akerberg et al. (2015).

levels as instruments, with a levels equation, using lagged differences as instruments. But rather than estimate the equations simultaneously, as in System GMM, we estimate them separately in what we call a two-step instrumental-variables (TSIV) procedure, in the spirit of a related exercise by Kripfganz and Schwarz (2019). In the first step, we first-difference and use the external instruments described above, together with a parsimonious set of internal instruments (lagged levels), to recover the coefficients on materials and labor, treating the capital coefficient as a nuisance parameter. In the second step, we use the first-step estimates of the materials and labor coefficients and impose an additional assumption that ensures orthogonality between the lagged difference in log capital and the firm effect; this allows us to use the lagged difference of capital as an instrument in an IV estimator in levels.⁷ If the model is correctly specified, the TSIV estimator is less efficient than simultaneous GMM estimation of the difference and levels equations, but it has the advantage that the materials and labor coefficient estimates are robust to possible misspecification of the levels equation (Kripfganz and Schwarz, 2019).

The TSIV procedure yields plausible point estimates: we find materials and labor coefficients of 0.45 and 0.47, respectively, and a capital coefficient of 0.11. The fact that constant returns to scale approximately hold, as one would generally expect (Bartelsman and Doms, 2000), is reassuring. Although our confidence intervals are large enough that the differences with standard estimators are generally not statistically significant, the point estimates display some interesting patterns. Compared to the Olley and Pakes (1996) and Levinsohn and Petrin (2003) proxy-variable methods, we find a smaller materials coefficient and larger labor coefficient. Compared to System GMM, we find a larger labor coefficient. Our estimates are most similar to those of Gandhi et al. (2020). To further investigate the performance of our estimator relative to other common estimators, we conduct a simple Monte Carlo simulation, considering a series of data-generating processes (DGPs) that are consistent with our theoretical framework. We abstract from variety and output-quality differences and examine the roles of imperfect output-market competition, firm fixed effects, and idiosyncratic input-quality shocks. We find that these features adversely affect other common estimators but that our estimator continues to perform well.

In addition to the studies cited above, this paper is related to several branches of literature. It is perhaps most closely related to a small number of studies on production-function estimation in multi-product firms using information at the firm-product level. This literature has dealt in different ways with the lack of an observed mapping between inputs and outputs in multi-product firms. One strategy has been to focus on single-product firms, for which the mapping is clear, and (in some cases) to do a selection correction for the fact that they may not be representative (Foster et al., 2008; De Loecker et al., 2016; Garcia-Marin and Voigtländer, 2019; Blum et al., forthcoming; Forlani et al., forthcoming).⁸ Dhyne et al. (2022, 2023) develop an alternative strategy that can

⁷The second-step approach is akin to that of Collard-Wexler and De Loecker (2020).

⁸Foster et al. (2008) include in their sample only firms in which one product makes up more than 50% of revenues,

be applied in multi-product firms, in which they relate output of a good to firm-level input usage and the output levels of other goods. Another approach has been to use estimates of demand elasticities and profit-maximization conditions to infer the allocation of inputs to outputs that would be implemented by optimizing firms (Gong and Sickles, 2021; Orr, 2022; Valmari, forthcoming). Our strategy, by contrast, is to aggregate both outputs and inputs to the firm level. Previous papers that have aggregated from the firm-product to the firm level, without explicitly considering quality and variety biases, include Eslava et al. (2004, 2013), Ornaghi (2006), Doraszelski and Jaumandreu (2013), Smeets and Warzynski (2013), and Garcia-Marin and Voigtländer (2019). Papers that have used CES aggregation in a production-function context include Halpern et al. (2015), Doraszelski and Jaumandreu (2018), and Harrigan et al. (2021). Our approach builds on an extensive literature using CES functions in addressing other questions, including Feenstra (1994), Hottman et al. (2016), and Redding and Weinstein (2020).

This paper is also related to studies that explicitly consider differences in the quality of outputs or inputs in a production-function context. Melitz (2000), Katayama et al. (2009), and Grieco et al. (2016) propose estimators that take quality differences into account in settings where product-level information is not observed; the lack of direct price and quantity data means that they must rely on more restrictive theoretical assumptions than we do here. Fox and Smeets (2011) show that including detailed indicators of labor quality significantly reduces the dispersion of estimated productivities across firms in Denmark, but they do not have product-level information on outputs or material inputs. For the most part, the literature exploiting information at the firm-product level does not explicitly take into account quality or variety differences. Exceptions include De Loecker et al. (2016), who use a control-function approach to capture input quality differences, and Eslava et al. (forthcoming), who use quality-adjusted deflators constructed via joint estimation of production functions and demand.⁹ Two recent papers take advantage of detailed product characteristics in particular sectors. Focusing on outpatient dialysis centers in the US, Grieco and McDevitt (2016) find that firms trade off quality and quantity of care, suggesting that measures of performance based solely on quantity can be misleading. In an Egyptian rug cluster, Atkin et al. (2019) collect direct measures of rug quality and producer performance under laboratory conditions and also find that purely quantity-based measures of performance are misleading. Such direct measures of product quality are clearly very valuable for estimating firm performance, but unfortunately they are rarely

essentially focusing on single-product firms. De Loecker et al. (2016) implement a modified version of the Akerberg et al. (2015) proxy-variable method, focusing first on single-product firms and then doing a selection correction to calculate firm-product-level markups in Indian data. Garcia-Marin and Voigtländer (2019) calculate firm-product-level markups along the same lines in Chilean data.

⁹De Loecker et al. (2016) put flexible functions of output prices and market shares on the right-hand side and physical quantities of output on the left-hand side. This approach arguably removes quality biases in the special case where input and output quality are perfectly correlated, but does not address what we call variety biases or the more general case where input and output quality are not perfectly correlated. Eslava et al. (forthcoming) (contemporaneously with this paper) also use CES aggregation, but in the context of joint GMM estimation of production and demand functions that requires CES across as well as within firms.

available. We view our method as being most useful in settings where product prices and quantities are available but detailed product characteristics are not.

Relative to the existing literature, our approach has benefits and costs. We are able to avoid some strong assumptions required by other methods. We do not need a scalar monotonicity condition to ensure invertibility of an investment or materials-demand function as in standard proxy-variable methods. This is particularly convenient in settings where productivity or quality (in inputs or outputs) have important firm fixed components. We do not need first-order conditions for aggregate materials or labor to hold exactly as in Doraszelski and Jaumandreu (2013, 2018), and Gandhi et al. (2020), and Demirer (2022). We can remain agnostic about cross-firm demand elasticities. Relative to the panel-data literature, we are able to reduce the reliance on lagged internal instruments. If we are only interested in the output elasticities with respect to materials and labor, we can avoid the assumptions required for the levels equation in System GMM. On the other hand, some of our assumptions are stronger than those of other methods. For our internal instruments to be valid, we require unusually strong, but testable, restrictions on the evolution of unobserved productivity, prices, quality and variety. For our external instruments, we require exclusion restrictions (discussed in detail below). We do not consider firms' endogenous investments in raising productivity, as do for instance Doraszelski and Jaumandreu (2013, 2018). The within-firm CES assumptions are restrictive (although the empirical patterns are robust to using other aggregators). While there are trade-offs, we believe that, on balance, our method represents an attractive alternative to existing methods in differentiated-product industries where quantity information and external instruments are available.

The next section develops our econometric strategy. Section 3 describes the data we use and our motivation for focusing on producers of rubber and plastic products. Section 4 presents our estimates of output elasticities. Section 5 compares our coefficient estimates to those of other common estimation methods, first in the Colombian data (Subsection 5.1) and then in a Monte Carlo simulation (Subsection 5.2). Section 6 concludes.

2 Econometric Strategy

This section first lays out the CES decompositions on the demand side (Subsection 2.1) and production side (Subsection 2.2) and then uses the decompositions to rewrite the production function, which makes clear how endogenous quality and variety choices may bias standard estimates (Subsection 2.3). We then discuss the timing (and other) assumptions we impose (Subsection 2.4) and present our two-step IV (TSIV) strategy (Subsection 2.5). Full derivations are in Appendix A.

2.1 Demand: Set-up and CES Decomposition

The first task is to construct a measure of real output at the firm level — firm-level sales deflated by an appropriate firm-specific price index. In differentiated-good industries, any price index necessarily

embeds assumptions about how a firm’s products enter consumers’ utility. Here we follow Hottman et al. (2016) and others in imposing constant elasticity of substitution of products within firms. This is restrictive, but unlike much of the existing literature we do not need to make strong assumptions about the elasticity of substitution of products across firms. (We will also show in Appendix C.1 that the empirical patterns are robust to using other common aggregators.)

We assume that a representative consumer has the following utility function:

$$U_t = U(\tilde{Y}_{1t}, \tilde{Y}_{2t}, \dots, \tilde{Y}_{It}) \quad \text{where} \quad \tilde{Y}_{it} = \left[\sum_{j \in \Omega_{it}^y} (\varphi_{ijt} Y_{ijt})^{\frac{\sigma_i^y - 1}{\sigma_i^y}} \right]^{\frac{\sigma_i^y}{\sigma_i^y - 1}} \quad (1)$$

where $U(\cdot)$ is quasi-concave and weakly separable in its arguments. Here i , j and t index firms, products (outputs), and periods (years), I is the total number of firms, Y_{ijt} is physical quantity of output, σ_i^y is the elasticity of substitution between outputs, specific to firm i , and Ω_{it}^y is the set of products sold by the firm. The φ_{ijt} terms are demand shifters that can be interpreted as product quality (or “appeal”), which may reflect endogenous choices of the firm (e.g. physical attributes of goods) or external factors (e.g. exogenous fashion trends). We follow common practice and assume that $\sigma_i^y > 1$.¹⁰

The assumption of weak separability and the homotheticity of \tilde{Y}_{it} imply that the consumer’s optimization problem can be solved in two stages, first choosing the quantity of each variety from firm i , Y_{ijt} , to minimize the cost of acquiring each unit of \tilde{Y}_{it} and then choosing \tilde{Y}_{it} to maximize utility. Optimization in the first stage implies that the minimum price to purchase one unit of \tilde{Y}_{it} is:

$$\tilde{P}_{it} = \left[\sum_{j \in \Omega_{it}^y} \left(\frac{P_{ijt}}{\varphi_{ijt}} \right)^{1 - \sigma_i^y} \right]^{\frac{1}{1 - \sigma_i^y}} \quad (2)$$

This is the price index that sets $\tilde{P}_{it} \tilde{Y}_{it} = R_{it}$, where R_{it} is the consumer’s total expenditures on goods of firm i , which are also the firm’s revenues. Note that the price index is quality-adjusted: conditional on prices, higher output quality reduces the value of the index.

An attractive feature of our approach is that we do not need to impose further assumptions on demand. The assumption of quasi-concavity implies that there is a unique demanded bundle, given by:

$$\tilde{Y}_{it} = D_{it}(\tilde{P}_{1t}, \dots, \tilde{P}_{It}, C_t) \quad \text{for } i = 1, 2, \dots, I \quad (3)$$

¹⁰Although the consumer optimization problem would remain well-behaved as long as $\sigma_i^y \in (0, 1) \cup (1, \infty)$ (see Appendix A.1), the stronger $\sigma_i^y > 1$ ensures that the representative consumer will purchase more units of a good that increases in appeal, which seems realistic in our context. As noted by Redding and Weinstein (2020), $\sigma_i^y > 1$ is sufficient to ensure that products are “connected substitutes” in the sense of Berry et al. (2013) and hence that the demand system is invertible.

where C_t is total consumption in period t . The demand for the output aggregate of a given firm depends only on the firm's own aggregate price index, the price indexes of other firms, and total consumption. We do not need to restrict the $D(\cdot)$ function further.

The within-firm CES assumption allows us to decompose changes in the firm-specific price index in a particularly convenient way. Let $\Omega_{it,t-1}^{y*}$ be firm i 's common outputs between $t-1$ and t (i.e. $\Omega_{it-1}^y \cap \Omega_{it}^y$), S_{ijt}^y be the consumer's period- t expenditure share on product j among all products produced by firm i , $S_{ijt,t-1}^{y*}$ be the period- t share among $(t, t-1)$ common goods, and $S_{ijt-1,t}^{y*}$ be the period- $(t-1)$ share among $(t, t-1)$ common goods.¹¹ Following Feenstra (1994) and Redding and Weinstein (2020), it is straightforward to show (see Appendix A.1) that the log change in the firm-specific price level can be expressed as:

$$\ln\left(\frac{\tilde{P}_{it}}{\tilde{P}_{it-1}}\right) = \sum_{j \in \Omega_{it,t-1}^{y*}} \psi_{ijt}^y \ln\left(\frac{P_{ijt}}{P_{ijt-1}}\right) - \sum_{j \in \Omega_{it,t-1}^{y*}} \psi_{ijt}^y \ln\left(\frac{\varphi_{ijt}}{\varphi_{ijt-1}}\right) - \frac{1}{\sigma_i^y - 1} \ln\left(\frac{\chi_{it-1,t}^y}{\chi_{it,t-1}^y}\right) \quad (4)$$

where:

$$\psi_{ijt}^y = \frac{\left(\frac{S_{ijt,t-1}^{y*} - S_{ijt-1,t}^{y*}}{\ln S_{ijt,t-1}^{y*} - \ln S_{ijt-1,t}^{y*}}\right)}{\sum_{j \in \Omega_{it,t-1}^{y*}} \left(\frac{S_{ijt,t-1}^{y*} - S_{ijt-1,t}^{y*}}{\ln S_{ijt,t-1}^{y*} - \ln S_{ijt-1,t}^{y*}}\right)}, \quad \chi_{it,t-1}^y = \sum_{j \in \Omega_{it,t-1}^{y*}} S_{ijt}^y, \quad \chi_{it-1,t}^y = \sum_{j \in \Omega_{it,t-1}^{y*}} S_{ijt-1}^y \quad (5)$$

The first term on the right-hand side of (4) is (the log of) the familiar Sato-Vartia index; it is an weighted average of product-specific price changes for common goods, with the ‘‘Sato-Vartia weights’’ ψ_{ijt}^y ; note that it is observable in data like ours. The second term is a weighted average of changes in (unobservable) product quality, again using the Sato-Vartia weights. Intuitively, increases in product quality reduce the price index, other things equal.¹² The third term is an adjustment for entry and exit of products, first introduced by Feenstra (1994). Increases in product variety also tend to reduce the price index.¹³ Although the σ_i^y term is unobservable, the $\chi_{it-1,t}^y$ and $\chi_{it,t-1}^y$ terms (which capture the $(t, t-1)$ common-goods shares of total firm revenues in periods $t-1$ and t) are observable in our data.

Appendix A.1 further shows that the log change in the quantity index, \tilde{Y}_{it} , can be expressed in

¹¹That is, $S_{ijt}^y = \frac{P_{ijt}Y_{ijt}}{R_{it}}$, and, for $j \in \Omega_{it,t-1}^{y*}$, $S_{ijt,t-1}^{y*} = \frac{P_{ijt}Y_{ijt}}{\sum_{j' \in \Omega_{it,t-1}^{y*}} P_{ij't}Y_{ij't}}$ and $S_{ijt-1,t}^{y*} = \frac{P_{ijt-1}Y_{ijt-1}}{\sum_{j' \in \Omega_{it,t-1}^{y*}} P_{ij't-1}Y_{ij't-1}}$.

¹²Redding and Weinstein (2020), in a very different exercise, deal with the quality terms by assuming that the geometric average of product quality across products is time-invariant; our approach, by contrast, is to assume that they are orthogonal to the instruments we construct, as will be made clear below.

¹³For example, if no goods are dropped from $t-1$ to t but new goods are introduced, then $\chi_{it-1,t}^y = 1 > \chi_{it,t-1}^y$, which, since $\sigma_i^y > 1$ by assumption, implies a reduction in the price index. This reflects the fact that the utility function (1) embeds a taste for variety in the goods from a given firm.

a similar decomposition:

$$\ln\left(\frac{\tilde{Y}_{it}}{\tilde{Y}_{it-1}}\right) = \sum_{j \in \Omega_{it,t-1}^{y*}} \psi_{ijt}^y \ln\left(\frac{Y_{ijt}}{Y_{ijt-1}}\right) + \sum_{j \in \Omega_{it,t-1}^{y*}} \psi_{ijt}^y \ln\left(\frac{\varphi_{ijt}}{\varphi_{ijt-1}}\right) + \frac{\sigma_i^y}{\sigma_i^y - 1} \ln\left(\frac{\chi_{it-1,t}^y}{\chi_{it,t-1}^y}\right) \quad (6)$$

The first term is again the log of a Sato-Vartia index, this time for quantities, the second term captures improvements in product quality, and the third term captures increases in product variety.

It is worth noting that this within-firm CES approach nests the common approach of using firm revenues deflated by a sector-level price index to measure real output, as $\sigma_i^y \rightarrow \infty$.¹⁴ In that sense, our aggregation method is strictly more general than the common approach.

2.2 Production: Set-up and CES Decomposition

On the production side, we assume that real output, as defined above, is produced as a function of capital, labor, and a firm-level CES materials aggregate, combining in Cobb-Douglas fashion:

$$\tilde{Y}_{it} = \tilde{M}_{it}^{\beta_m} L_{it}^{\beta_\ell} K_{it}^{\beta_k} e^{\omega_{it} + \eta_i + \xi_t + \epsilon_{it}} \quad \text{where} \quad \tilde{M}_{it} = \left[\sum_{h \in \Omega_{it}^m} (\alpha_{iht} M_{iht})^{\frac{\sigma_i^m - 1}{\sigma_i^m}} \right]^{\frac{\sigma_i^m}{\sigma_i^m - 1}} \quad (7)$$

Here h indexes material inputs, Ω_{it}^m is the set of inputs purchased by the firm, M_{iht} is the quantity of each material input purchased, L_{it} is labor, and K_{it} is capital. We refer to α_{iht} as input quality, recognizing that it may reflect physical attributes of the inputs or characteristics of the technology used to combine them in production. The $\{\alpha_{iht}\}$ capture any differences across inputs in how much one physical unit of an input contributes to the input aggregate. The assumption that the production function is Cobb-Douglas in capital, labor, and materials is standard in the literature. In principle, our approach could be extended to other functional forms (e.g. translog), although other forms would most likely require additional instruments. As on the output side, we assume the the firm-specific elasticity of substitution between inputs is greater than unity, $\sigma_i^m > 1$, which ensures that a firm consumes more of an input that increases in quality. In addition to being standard, this assumption is consistent with recent evidence at the micro level that intermediate inputs are typically substitutes (Dhyne et al., 2022; Peter and Ruane, 2020).¹⁵

The terms ω_{it} , η_i , ξ_t , and ϵ_{it} reflect firm productivity, where productivity should be understood as the capability, given input levels, to produce *real output*, i.e. the CES bundle \tilde{Y}_{it} , which incorporates quality and variety as well as physical units. The ω_{it} term is an “ex ante” shock that firms observe

¹⁴From (1) and (2), $\lim_{\sigma_i^y \rightarrow \infty} \tilde{Y}_{it} = \sum_{j \in \Omega_{it}^y} \varphi_{ijt} Y_{ijt}$ and $\lim_{\sigma_i^y \rightarrow \infty} \tilde{P}_{it} = \min_{j \in \Omega_{it}^y} (P_{ijt}/\varphi_{ijt})$. In this case, all goods purchased by the consumer have the same quality-adjusted price, call it $\check{P}_t = P_{ijt}/\varphi_{ijt} \forall j \in \Omega_{it}^y$; goods with higher quality-adjusted prices are not purchased. Then $R_{it} = \sum_{j \in \Omega_{it}^y} P_{ijt} Y_{ijt} = \sum_{j \in \Omega_{it}^y} (P_{ijt}/\varphi_{ijt}) \varphi_{ijt} Y_{ijt} = \check{P}_t \tilde{Y}_{it}$. Hence as $\sigma_i^y \rightarrow \infty$, deflating R_{it} by \check{P}_t yields real output.

¹⁵As on the output side (see footnote 10), our method remains applicable, although with somewhat less intuitive implications, as long as $\sigma_i^m \in (0, 1) \cup (1, \infty)$.

before choosing flexible inputs. The η_i term is a time-invariant firm effect. The ξ_t term is a sector- or economy-level shock. The ϵ_{it} term is an “ex post” shock that is only revealed to firms after they have chosen inputs (and hence is not “transmitted” to input choices); it may also capture measurement error.

We allow input and output variety and quality to be chosen endogenously by firms. Researchers have proposed a number of models for such choices; see for instance Eckel and Neary (2010) and Bernard et al. (2011) on variety, and Kugler and Verhoogen (2012) on quality. Here we do not adopt a particular model of how firms make these choices. We discuss the assumptions we need on variety and quality choices in Section 2.4 below. We believe that it is most natural to think of firms as first choosing variety and quality and then choosing values of L_{it} and $\{M_{iht}\}$ (all within period t); we proceed under that assumption.

Conditional on choices of input and output quality and variety, the derivations of the price and quantity indexes on the input side are analogous to those on the output side. Given the production function, (7) (which is also weakly separable, with homothetic aggregate \widetilde{M}_{it}), the firm can be thought of as first choosing input quantities, M_{iht} , to minimize the cost of acquiring a given level of the aggregate input, \widetilde{M}_{it} , and then choosing optimal values of \widetilde{M}_{it} , L_{it} and investment in capital, given the demand function, (3). Optimization in the choice of input quantities implies that the cost of purchasing one unit of the materials aggregate, \widetilde{M}_{it} , is:

$$\widetilde{W}_{it}^m = \left[\sum_{h \in \Omega_{it}^m} \left(\frac{W_{iht}^m}{\alpha_{iht}} \right)^{1-\sigma_i^m} \right]^{\frac{1}{1-\sigma_i^m}} \quad (8)$$

This is the price index that sets $\widetilde{W}_{it}^m \widetilde{M}_{it} = E_{it}^m$, where E_{it}^m is the firm’s total expenditures on material inputs.

As on the output side, we can decompose input-price changes in a convenient way. Let $\Omega_{it,t-1}^{m*}$ be firm i ’s common inputs between $t-1$ and t (i.e. $\Omega_{it-1}^m \cap \Omega_{it}^m$), $E_{it,t-1}^{m*}$ be the firm’s expenditures on $(t, t-1)$ common inputs, $S_{iht,t-1}^m$ be the firm’s expenditure share on input h among all inputs purchased by firm i , $S_{iht,t-1}^{m*}$ be the period- t share among $(t, t-1)$ common inputs, and $S_{iht,t-1}^{m*}$ be the period- $(t-1)$ share among $(t, t-1)$ common inputs.¹⁶ The log change in the firm-specific input price level can be expressed as:

$$\ln \left(\frac{\widetilde{W}_{it}^m}{\widetilde{W}_{it-1}^m} \right) = \sum_{h \in \Omega_{it,t-1}^{m*}} \psi_{iht}^m \ln \left(\frac{W_{iht}^m}{W_{iht-1}^m} \right) - \sum_{h \in \Omega_{it,t-1}^{m*}} \psi_{iht}^m \ln \left(\frac{\alpha_{iht}}{\alpha_{iht-1}} \right) - \frac{1}{\sigma_i^m - 1} \ln \left(\frac{\chi_{it-1,t}^m}{\chi_{it,t-1}^m} \right) \quad (9)$$

¹⁶That is, $S_{iht}^m = \frac{W_{iht}^m M_{iht}}{E_{it}^m}$, and, for $h \in \Omega_{it,t-1}^{m*}$, $S_{iht,t-1}^{m*} = \frac{W_{iht}^m M_{iht}}{\sum_{h' \in \Omega_{it,t-1}^{m*}} W_{ih't}^m M_{ih't}}$ and $S_{iht-1,t}^{m*} = \frac{W_{iht-1}^m M_{iht-1}}{\sum_{h' \in \Omega_{it,t-1}^{m*}} W_{ih't-1}^m M_{ih't-1}}$.

where:

$$\psi_{iht}^m = \frac{\left(\frac{S_{iht,t-1}^{m*} - S_{iht-1,t}^{m*}}{\ln S_{iht,t-1}^{m*} - \ln S_{iht-1,t}^{m*}} \right)}{\sum_{h \in \Omega_{it,t-1}^{m*}} \left(\frac{S_{iht,t-1}^{m*} - S_{iht-1,t}^{m*}}{\ln S_{iht,t-1}^{m*} - \ln S_{iht-1,t}^{m*}} \right)}, \quad \chi_{it,t-1}^m = \sum_{h \in \Omega_{it,t-1}^{m*}} S_{iht}^m, \quad \chi_{it-1,t}^m = \sum_{h \in \Omega_{it,t-1}^{m*}} S_{iht-1}^m \quad (10)$$

As for output prices, the first term is the observable log Sato-Vartia price change index for common goods, the second term is a weighted average of changes in input quality, and the third term is an adjustment for entry and exit of inputs.

As for output quantities, the change in the CES materials quantity aggregate can be written as the sum of an observable Sato-Vartia quantity change index and unobservable terms capturing increases in quality and variety:¹⁷

$$\ln \left(\frac{\widetilde{M}_{it}}{\widetilde{M}_{it-1}} \right) = \sum_{h \in \Omega_{it,t-1}^{m*}} \psi_{iht}^m \ln \left(\frac{M_{iht}}{M_{iht-1}} \right) + \sum_{h \in \Omega_{it,t-1}^{m*}} \psi_{iht}^m \ln \left(\frac{\alpha_{iht}}{\alpha_{iht-1}} \right) + \frac{\sigma_i^m}{\sigma_i^m - 1} \ln \left(\frac{\chi_{it-1,t}^m}{\chi_{it,t-1}^m} \right) \quad (11)$$

2.3 Estimating Equations

To integrate the CES output and input quantity decompositions (6) and (11) into the production function, (7), it is convenient to restate the decompositions in levels. Let lower-case letters represent logs and Δ a change from $t-1$ to t . Summing the differences in (6) and (11) over time within firms, with firm-specific normalizations \widetilde{y}_{i0} and \widetilde{m}_{i0} , we have:

$$\begin{aligned} \widetilde{y}_{it} &= \underbrace{\widetilde{y}_{i0} + \sum_{\tau=1}^t \sum_{j \in \Omega_{i\tau,\tau-1}^{y*}} \psi_{ij\tau}^y \Delta y_{ij\tau}}_{=\widetilde{y}_{it}^{SV}} + \underbrace{\sum_{\tau=1}^t \sum_{j \in \Omega_{i\tau,\tau-1}^{y*}} \psi_{ij\tau}^y \ln \left(\frac{\varphi_{ij\tau}}{\varphi_{ij\tau-1}} \right)}_{=q_{it}^y} + \underbrace{\left(\frac{\sigma_i^y}{\sigma_i^y - 1} \right) \sum_{\tau=1}^t \ln \left(\frac{\chi_{i\tau-1,\tau}^y}{\chi_{i\tau,\tau-1}^y} \right)}_{=v_{it}^y} \quad (12) \\ \widetilde{m}_{it} &= \underbrace{\widetilde{m}_{i0} + \sum_{\tau=1}^t \sum_{h \in \Omega_{i\tau,\tau-1}^{m*}} \psi_{ih\tau}^m \Delta m_{ih\tau}}_{=\widetilde{m}_{it}^{SV}} + \underbrace{\sum_{\tau=1}^t \sum_{h \in \Omega_{i\tau,\tau-1}^{m*}} \psi_{ih\tau}^m \ln \left(\frac{\alpha_{ih\tau}}{\alpha_{ih\tau-1}} \right)}_{=q_{it}^m} + \underbrace{\left(\frac{\sigma_i^m}{\sigma_i^m - 1} \right) \sum_{\tau=1}^t \ln \left(\frac{\chi_{i\tau-1,\tau}^m}{\chi_{i\tau,\tau-1}^m} \right)}_{=v_{it}^m} \end{aligned}$$

where we define \widetilde{y}_{it}^{SV} , q_{it}^y , v_{it}^y , \widetilde{m}_{it}^{SV} , q_{it}^m , and v_{it}^m , as indicated by the underbraces. In defining variables in this way, we are setting the quality and variety terms q_{it}^y , v_{it}^y , q_{it}^m , and v_{it}^m to zero in the initial year and including the firm-specific normalizations as part of the Sato-Vartia quantity terms, \widetilde{y}_{it}^{SV} and \widetilde{m}_{it}^{SV} . (Although these normalizations will be differenced out in first-differences, they will become relevant in the second step of our two-step IV estimation below, in levels.)

¹⁷This approach again nests the standard approach of using expenditures deflated by a sector-level input price index as $\sigma_i^m \rightarrow \infty$; see footnote 14.

Plugging these expressions into the production function, (7), and rearranging, we have:

$$\begin{aligned} \tilde{y}_{it}^{SV} &= \beta_m \tilde{m}_{it}^{SV} + \beta_\ell \ell_{it} + \beta_k k_{it} + \eta_i + \xi_t + u_{it}, \\ \text{where } u_{it} &= (\beta_m v_{it}^m - v_{it}^y) + (\beta_m q_{it}^m - q_{it}^y) + \omega_{it} + \epsilon_{it} \end{aligned} \quad (13)$$

This equation relates the Sato-Vartia output quantity index to the Sato-Vartia input quantity index (both observable), log capital, log labor, a firm effect, a year effect, and an error term that reflects variety and quality of outputs and inputs as well as the “ex ante” and “ex post” productivity shocks. We refer to (13) as our levels equation.

Writing the production function in this way helps to clarify two issues. The first is that simply using physical quantities for output and input may be problematic in a setting where quality or variety vary differently by firm over time, on the output side or the input side. The input choices \tilde{m}_{it}^{SV} , ℓ_{it} , and k_{it} may be correlated with the unobserved quality and variety terms, q_{it}^m , q_{it}^y , v_{it}^m , and v_{it}^y , generating what we call *input- or output-quality biases*, or *input- or output-variety biases*. To fix ideas, suppose that firms produce a single product using a single material input, in which case \tilde{y}_{it}^{SV} and \tilde{m}_{it}^{SV} simplify to the physical quantities of the output and the input and the variety terms drop out. If producing one unit of a higher-quality output requires more physical units of labor, then there will be a positive correlation between ℓ_{it} and the q_{it}^y and hence a negative output-quality bias in the OLS estimate of β_ℓ . Biases may also arise from purely exogenous shocks to product appeal or input quality, if such shocks affect firms’ input choices — for instance, if a firm’s product becomes fashionable for exogenous reasons and the firm expands production to take advantage of the increased demand, or if a supplier improves the quality of a purchased input without changing the price and this induces the firm to increase output. Among multi-product, multi-input firms, biases can arise from changes in variety. For instance, if import-tariff reductions increase the set of input varieties available and induce firms to increase the variety of inputs purchased, the variety of outputs produced, and total output, as suggested by Goldberg et al. (2010), one would expect positive correlations between \tilde{m}_{it}^{SV} and v_{it}^m and between \tilde{m}_{it}^{SV} and v_{it}^y , generating offsetting biases with ambiguous net effects. It is important to note that these quality and variety biases are distinct from transmission bias, *and might be present even if one had a perfect proxy for the ex ante productivity term*, ω_{it} .

The second issue that equation (13) clarifies is why the scalar monotonicity assumption required by standard proxy-variable approaches is incompatible with our approach to aggregation. The leading proxy-variable approaches require a one-to-one relationship between a firm’s underlying productivity and either investment or materials demand, conditional on other observables (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; De Loecker, 2011; Doraszelski and Jaumandreu, 2013, 2018; Akerberg et al., 2015) As noted by Akerberg et al. (2015), in models with a firm effect (here η_i) in addition to the ex ante productivity term (here ω_{it}), this assumption is unlikely to hold, since the firm effect

introduces a second dimension of heterogeneity between firms.¹⁸ In our case, we assumed the presence of a firm effect at the outset, in the production function, (7). But even if we had not, we would have to deal with the firm-specific normalizations \tilde{y}_{i0} and \tilde{m}_{i0} in (12), which we have folded into the levels of the observable Sato-Vartia quantity aggregates, \tilde{y}_{it}^{SV} and \tilde{m}_{it}^{SV} . We impose a particular normalization in the second step of our estimation procedure below, but we feel that a strength of our approach is that we do not need to impose such an assumption in the first step when estimating the coefficients on materials and labor. We pursue an approach more in the spirit of the panel-data literature, in part because it can more easily accommodate the fixed effect.

As noted above, we will estimate an equation in first-differences as well as in levels.¹⁹ In differences, (13) becomes:

$$\begin{aligned} \Delta \tilde{y}_{it}^{SV} &= \beta_m \Delta \tilde{m}_{it}^{SV} + \beta_\ell \Delta \ell_{it} + \beta_k \Delta k_{it} + \Delta \xi_t + \Delta u_{it}, \\ \text{where } \Delta u_{it} &= (\beta_m \Delta v_{it}^m - \Delta v_{it}^y) + (\beta_m \Delta q_{it}^m - \Delta q_{it}^y) + \Delta \omega_{it} + \Delta \epsilon_{it} \end{aligned} \quad (14)$$

We refer to (14) as our difference equation. Note that, given the definitions of v_{it}^m , v_{it}^y , q_{it}^m , and q_{it}^y in (12), Δu_{it} can be rewritten as:

$$\begin{aligned} \Delta u_{it} &= \beta_m \frac{\sigma_i^m}{\sigma_i^m - 1} \ln \left(\frac{\chi_{it-1,t}^m}{\chi_{it,t-1}^m} \right) - \frac{\sigma_i^y}{\sigma_i^y - 1} \ln \left(\frac{\chi_{it-1,t}^y}{\chi_{it,t-1}^y} \right) \\ &\quad + \beta_m \sum_{h \in \Omega_{it,t-1}^{m*}} \psi_{iht}^m \ln \left(\frac{\alpha_{iht}}{\alpha_{iht-1}} \right) - \sum_{j \in \Omega_{it,t-1}^{y*}} \psi_{ijt}^y \ln \left(\frac{\varphi_{ijt}}{\varphi_{ijt-1}} \right) + \Delta \omega_{it} + \Delta \epsilon_{it} \end{aligned} \quad (15)$$

2.4 Timing (and Other) Assumptions

This section lays out the timing assumptions in our approach, under which our use of internal instruments is valid. Following Levinsohn and Petrin (2003) and others, we assume that materials and labor are flexible inputs, i.e. with no adjustment costs. We also make the standard assumption that capital can be adjusted only with a lag of one period.

Now consider the ex-ante and ex-post productivity shocks, ω_{it} and ϵ_{it} . Let \mathcal{I}_{it} be the information available to firm i when making production decisions in period t , which includes all past production choices, prices, realizations of ϵ_{is} and ω_{is} for all $s < t$, current-period log capital, k_{it} , the firm fixed

¹⁸This issue is separate from the concern that firms may face heterogeneous constraints in input markets, which might break the monotonic relationship between productivity and materials or investment demand (Shenoy, forthcoming). For recent efforts to accommodate firm fixed effects in proxy-variable frameworks, see Lee et al. (2019), Akerberg (2021), and Abito (2021).

¹⁹A “within” estimator, in which all variables are deviated from firm-specific means, would require the time-varying firm-specific productivity terms, ω_{it} and ϵ_{it} , to be uncorrelated with all past and future values of the covariates, which would be violated if productivity shocks affect future input choices. First-differencing yields an estimator that remains consistent under sequential exogeneity, where ω_{it} and ϵ_{it} are uncorrelated with past values of covariates but not necessarily future ones.

effect, η_i , the year effect, ξ_t , and the realization of ω_{it} (but not of ϵ_{it}) for period t . We assume:

$$\mathbb{E}(\epsilon_{it} | \mathcal{S}_{it}) = 0 \quad (16)$$

$$\mathbb{E}(\omega_{it} | \mathcal{S}_{it-1}) = 0 \quad (17)$$

Since $\epsilon_{it-1} \in \mathcal{S}_{it}$ and $\omega_{it-1} \in \mathcal{S}_{it-1}$, (16) and (17) imply that both ϵ_{it} and ω_{it} are serially uncorrelated, conditional on the firm fixed effect. The former assumption is standard. While the latter assumption is restrictive — the literature usually assumes a first-order Markov process in the ex-ante productivity term (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Akerberg et al., 2015; Gandhi et al., 2020) — we note that the firm fixed effect, η_i , is likely to capture the within-firm persistence that might show up as serial correlation in other approaches.

As noted above, we assume that variety and quality decisions are made before input quantity decisions. We allow input and output variety and quality to be chosen endogenously, but we assume that these choices depend only on contemporaneous variables. Formally, let $\bar{\Omega}_{it}^m$ and $\bar{\Omega}_{it}^y$ be vectors of 0/1 indicators for all possible inputs and outputs, where the 1's indicate the chosen inputs and outputs in Ω_{it}^m and Ω_{it}^y . Let $\bar{\alpha}_{it}$ and $\bar{\varphi}_{it}$ be vectors of qualities for inputs in Ω_{it}^m and outputs in Ω_{it}^y . Let $\bar{\Gamma}_{it}^{vm}$, $\bar{\Gamma}_{it}^{qm}$, $\bar{\Gamma}_{it}^{vy}$, and $\bar{\Gamma}_{it}^{qy}$ be vectors of exogenous, potentially firm-specific shifters that may affect input and output variety and quality. We assume that these shifters are serially uncorrelated.²⁰ Variety and quality choices are assumed to be determined as follows:

$$\bar{\Omega}_{it}^m = F^m(\omega_{it}, \eta_i, \bar{\Gamma}_{it}^{vm}) \quad (18a)$$

$$\bar{\Omega}_{it}^y = F^y(\omega_{it}, \eta_i, \bar{\Gamma}_{it}^{vy}) \quad (18b)$$

$$\bar{\alpha}_{it} = G^m(\omega_{it}, \eta_i, \bar{\Gamma}_{it}^{qm}) \quad (18c)$$

$$\bar{\varphi}_{it} = G^y(\omega_{it}, \eta_i, \bar{\Gamma}_{it}^{qy}) \quad (18d)$$

Given that ω_{it} and the exogenous shifters are serially uncorrelated by assumption, firms' variety and quality choices are also serially uncorrelated, conditional on η_i .

We assume that firms are price-takers in intermediate-input markets and that the input price for a given level of input quality can be written as the product of a firm-level average input price, \bar{W}_{it}^m , and a firm-input-specific term, ι_{iht} , which is serially uncorrelated:

$$W_{iht}^m = \bar{W}_{it}^m \iota_{iht} \quad \text{where } \mathbb{E}(\iota_{iht} | \mathcal{S}_{it-1}) = 0 \quad (19)$$

The key feature of (19) is that any serial correlation in input prices is limited to the firm-average price term, \bar{W}_{it}^m .²¹

²⁰Formally, $\mathbb{E}(\bar{\Gamma}_{it}^{vm} | \bar{\Gamma}_{it-1}^{vm}) = \mathbb{E}(\bar{\Gamma}_{it}^{vy} | \bar{\Gamma}_{it-1}^{vy}) = \mathbb{E}(\bar{\Gamma}_{it}^{qm} | \bar{\Gamma}_{it-1}^{qm}) = \mathbb{E}(\bar{\Gamma}_{it}^{qy} | \bar{\Gamma}_{it-1}^{qy}) = \bar{0}$.

²¹Note that, given the assumptions that materials and labor are flexible inputs and that the productivity shocks are serially uncorrelated, we will need \bar{W}_{it}^m to be serially correlated in order for lagged levels of materials and labor to have explanatory power for subsequent changes.

On the output side, the derivation of the times-series properties of the quality and variety terms, q_{it}^y and v_{it}^y , is complicated by the fact that firms endogeneously choose output prices, which in turn affect revenue shares (which enter the quality and variety terms). To solve explicitly for such output-price choices requires a specification of the micro-foundations for the firm-level production function, (7). A sufficient condition for our approach to be valid is that, like input prices, output prices can be written as the product of a potentially serially correlated firm-level term and a serially uncorrelated firm-product-level term:

$$P_{ijt} = \Lambda_{it}\varsigma_{ijt} \quad \text{where } \mathbb{E}(\varsigma_{ijt}|\mathcal{I}_{it}) = 0 \tag{20}$$

Appendix A.3 provides micro-foundations for our production function that are consistent with (20). Following Orr (2022), we assume that production functions across firm-products differ only in Hicks-neutral shifters and are homogeneous of degree one. We emphasize that these are not the only possible micro-foundations for our production function (7) and that our method will be valid as long as (20) holds.

Appendix A.4 shows that, under the assumptions laid out in this section, lagged levels from period $t-2$ or earlier are valid instruments. Intuitively, the firm-average prices in (19) and (20) drop out of the expressions for the input-expenditure shares, S_{iht}^m , and output-revenue shares, S_{ijt}^y , with the consequence (given our assumptions on quality and variety choices) that the shares are serially uncorrelated. Given the definitions of the Sato-Vartia weights, ψ_{iht}^m and ψ_{iht}^y , and the variety terms, $\chi_{it,t-1}^m$, $\chi_{it-1,t}^m$, $\chi_{it,t-1}^y$, and $\chi_{it-1,t}^y$ in (5) and (10), it follows from (15) that Δv_{it}^m , Δv_{it}^y , Δq_{it}^m , and Δq_{it}^y are $MA(1)$, i.e. display at most one period of serial correlation. Together with (16)-(17), this implies that input choices from $t-2$ and earlier are uncorrelated with Δu_{it} .

We recognize that the restrictions laid out in this section are stronger than those typically imposed in the literature. At the same time, a key implication of this set of assumptions — that Δu_{it} is $MA(1)$ — is testable using the standard approach of Arellano and Bond (1991). The results, report below, are consistent with this hypothesis: we will find strong correlation of Δu_{it} with Δu_{it-1} but we will not reject the hypothesis of no correlation between Δu_{it} and Δu_{it-2} . Serial correlation in the productivity shocks or in the determinants of quality or variety choices, conditional on the firm fixed effects, would be expected to show up as correlation between Δu_{it} and Δu_{it-2} (and further lags). The fact that we do not find such correlation increases our confidence in the set of assumptions we impose.

2.5 Two-Step IV Estimation Procedure

In the spirit of System GMM (Arellano and Bover, 1995; Blundell and Bond, 1998, 2000), we estimate both the difference and levels equations, but we proceed in two steps, along the lines of a related exercise in Kripfganz and Schwarz (2019). In the first step (Subsection 2.5.1), we estimate

the difference equation, (14), using lagged levels and external drivers of input price changes as instruments. We recover estimates of β_m and β_ℓ , treating β_k as a nuisance parameter. In the second step (Subsection 2.5.2), we insert the first-step estimates of β_m and β_ℓ in the levels equation and use the lagged difference of capital as an instrument for the level. To be clear on terminology, we use the word *step* to refer to the differences and levels steps, and reserve the word *stage* for the two stages of the IV estimator in each step.

2.5.1 Differences (Step 1)

To estimate the difference equation, (14), and to address quality and variety biases as well as the familiar transmission bias, we seek instruments that are correlated with $\Delta \tilde{m}_{it}^{SV}$, $\Delta \ell_{it}$ and Δk_{it} and uncorrelated with the error term, Δu_{it} . As noted above, lags of input choices from $t - 2$ and earlier are valid instruments under our assumptions. A widely recognized concern with lagged levels as instruments, however, is that they may be only weakly correlated with current differences (Griliches and Mairesse, 1998; Blundell and Bond, 1998, 2000; Bun and Windmeijer, 2010).²² Including further and further lags may exacerbate the weak-instrument problem. In our setting, diagnostic statistics will give reason to be concerned about the weakness of the internal instruments.

Our strategy for strengthening the instrument set is to incorporate external instruments capturing exogenous variation in the prices of materials and labor. To construct the materials-price instrument, we take advantage of detailed Colombian trade-transactions data merged with the Colombian manufacturing survey (for the pre-2009 period when this merge was possible). We first use real-exchange-rate movements to predict import-price movements at the product-year level, running “leave one out” regressions that omit one firm at a time.²³ We then use information on the product composition of each firm’s imports to aggregate the predictions to the firm-year level.

To be precise, we begin by defining real exchange rates (RERs) as:

$$RER_{ot} = NER_{ot} \left(\frac{CPI_{ot}}{CPI_{Col,t}} \right) \quad (21)$$

where o indexes import origins, NER_{ot} is the nominal exchange rate (Colombian pesos/foreign currency), CPI_{ot} is the consumer price index (CPI) in the origin, and $CPI_{Col,t}$ is the CPI in Colombia. Defined in this way, a real appreciation in country o is reflected in an increase in RER_{ot} . We consider the top 100 origins by Colombian import volume and label this set \mathcal{O} . We use n to index products defined at the 8-digit trade classification level (which do not map cleanly to products in the Colombian industrial classification, indexed by j and h above). We exclude machinery and equipment, which

²²In the setting of cross-country growth regressions, Bazzi and Clemens (2013) and Kraay (2015) show that the instruments used in difference and system GMM estimators are weak and can suggest misleading inferences. See also the review by Bun and Sarafidis (2015).

²³Exchange-rate movements have been used as a source of identification in similar contexts by Goldberg and Verboven (2001), Park et al. (2010), Bastos et al. (2018), Amiti et al. (2019), and others.

could arguably be considered capital rather than material imports; we also exclude petroleum and other mineral fuels.²⁴ For a particular imported input n , we calculate an average log RER change separately for each firm in our data, weighting by imports but leaving out the firm’s own imports:

$$\Delta \overline{rer}_{nt,-i} = \sum_{o \in \mathcal{O}} \zeta_{ont-1,-i} \Delta \ln(RER_{ot}), \quad \text{where } \zeta_{ont-1,-i} = \frac{\mathcal{I}_{ont-1,-i}}{\sum_{o \in \mathcal{O}} \mathcal{I}_{ont-1,-i}} \quad (22)$$

Here $\mathcal{I}_{ont-1,-i}$ is the “leave-one-out” value of imports of input n from origin o in period $t-1$ for all firms except i .²⁵ We then use these product-level average real-exchange-rate changes to predict import price changes at the product-year level, using the regression:

$$\Delta w_{nt,-i}^{imp} = \gamma_{st,-i} \Delta \overline{rer}_{nt,-i} + \rho_{st,-i} + \eta_{nt,-i} \quad (23)$$

where $\Delta w_{nt,-i}^{imp}$ is the change in import n ’s log import price (calculated at the product-year level, averaging across origins using import weights) for imports of all firms except i ,²⁶ $\rho_{st,-i}$ is a sector-year effect, and $\eta_{nt,-i}$ is a product-year-level disturbance. In our preferred specification, s indexes two digit trade sectors.²⁷ We run this leave-one-out regression separately for each firm i (using data from all firms present in both the customs data and the manufacturing survey, not just those in the rubber and plastics sectors) and recover the predicted values, $\Delta \widehat{w}_{nt,-i}^{imp}$. The advantage of using the predicted values, $\Delta \widehat{w}_{nt,-i}^{imp}$, as opposed to $\Delta w_{nt,-i}^{imp}$, is that the former reflect only the RER changes (and sector-year effects), which are credibly exogenous to firm i ’s decisions, and not shocks to quality or other unobserved characteristics of products (in $\eta_{nt,-i}$), which may be correlated across firms and hence potentially correlated with firm i ’s quality or variety choices, despite the fact that we have left out firm i in constructing $\Delta w_{nt,-i}^{imp}$.

We then use firm i ’s product-level import shares in $t-2$ as weights in constructing the average predicted import price change at the firm level:

$$\Delta \widehat{w}_{it}^{imp} = \sum_{n \in \mathcal{N}} \theta_{int-2} \Delta \widehat{w}_{nt,-i}^{imp}, \quad \text{where } \theta_{int-2} = \frac{\mathcal{I}_{int-2}}{\sum_{n \in \mathcal{N}} \mathcal{I}_{int-2}} \quad (24)$$

Here \mathcal{I}_{int-2} is imports by firm i of product n in period $t-2$ and \mathcal{N} is the set of all imported products. For firms that did not import in $t-2$, we set $\Delta \widehat{w}_{it}^{imp} = 0$.²⁸ This average predicted import price

²⁴That is, we exclude Harmonized System 2-digit categories 27, 84 and 85.

²⁵That is, $\mathcal{I}_{ont-1,-i} = \sum_{j \neq i} \mathcal{I}_{jnt-1}$ for input-year $nt-1$.

²⁶That is, $\Delta w_{nt,-i}^{imp} = \sum_{o \in \mathcal{O}} \zeta_{ont-1,-i} \Delta w_{ont,-i}^{imp}$, where $\zeta_{ont-1,-i}$ is defined as in (22).

²⁷In principle, we could include lags of the average real-exchange-rate changes in (23). But consistent with the literature on exchange-rate pass-through (see e.g. Campa and Goldberg (2005)), we have found that the effect of RER changes on import prices decays relatively quickly, within one year, and including further lags has little effect on the strength of our instrument, so we do not include them here.

²⁸If a concordance from detailed trade categories to detailed industrial categories were available, it would be possible to estimate the effects of RER changes on domestic prices and use firms’ composition of domestic purchases as well as imports to construct the firm-level instrument. But unfortunately no such concordance exists in Colombia. We experimented with constructing our own concordance based on verbal product descriptions, but we found this to be

change at the firm level, $\Delta \widehat{w}_{it}^{imp}$, is our external instrument for $\Delta \widetilde{m}_{it}^{SV}$ in (14).²⁹

In order for this instrument to be valid, it must be both correlated with input choices (which we will provide evidence of below) and uncorrelated with the error term in the difference equation, which appears in (14) and (15). Given that the instrument is a function of RER changes and firms' initial import compositions, it seems clearly uncorrelated with the productivity shocks, $\Delta \omega_{it}$ and $\Delta \epsilon_{it}$ in Δu_{it} . One might be more worried that it is correlated with the input and output quality and variety terms, for instance if the quality of imports varies systematically by origin and RER movements lead firms to source from different origins. Our main response to this concern, which we return to below, is that we have chosen to focus on sectors — plastic and rubber products — in which inputs are commodities or at least commodity-like. Although inputs such as natural latex and carbon blacks (for rubber products) or polyethylene and polypropylene (for plastic products) may have quality differences, they remain highly substitutable within quality categories observable to market participants and we would not expect significant differences in quality or variety across origin countries. We maintain the assumption that $\Delta \widehat{w}_{it}^{imp}$ is uncorrelated with Δu_{it} in what follows.

To construct an external instrument for labor, we exploit the fact that the minimum wage in Colombia is high relative to the wage distribution (above 90% of the median) and that it rose sharply over our sample period, especially in 1994-1999 and 2003-2009. (See Section 3.3 for institutional background.) We first construct a measure of the “bite” of the minimum wage — how binding it is expected to be on a particular firm — defined as:

$$B_{it} = \frac{MW_t}{W_{it}^\ell} \quad (25)$$

where MW_t^m is the national minimum wage (defined for monthly earnings and annualized by multiplying by 12) and W_{it}^ℓ is firm-level average annual earnings per worker for permanent workers, calculated as the firm-level annual wage bill divided by average employment. Defined in this way, $B_{it} < 1$; the closer the firm average wage is to the national minimum wage, the larger is B_{it} . We interact this measure of bite with the change in the national minimum wage, using bite from $t - 2$:

$$\Delta z_{it} = B_{it-2} * \Delta \ln(MW_t) \quad (26)$$

This predicted wage change, Δz_{it} , serves as an instrument for $\Delta \ell_{it}$ in (14). We maintain the assumption that Δz_{it} is uncorrelated with the differenced error term, Δu_{it} , conditional on the (differenced) year effect, $\Delta \xi_t$, and hence is a valid instrument. Previous studies that have followed this strategy of interacting minimum wage changes with differences in their bite include Card (1992) and Cengiz et al. (2019).

impracticable.

²⁹Exchange-rate movements may also affect export prices. We address this concern by constructing an analogous predicted export price index and including it as an additional covariate; see Section C.2 below.

It is well known that the estimation of the capital coefficient is problematic in models that include a firm effect. For example, in a first-differenced model using lagged levels as instruments, Ornaghi (2006) finds a negative coefficient on capital. Using a within estimator, Söderbom and Teal (2004) also find a negative relationship. It is common to attribute low estimates of the capital coefficient to measurement error in capital, the effect of which is exacerbated by transformations to remove the firm effect (Griliches and Mairesse, 1998; Akerberg et al., 2015). In the Colombian manufacturing census, we do not observe capital utilization, and it seems likely that the capital measure we are able to construct, while standard, is a very noisy measure of capital actually in use. It may also be that in the presence of adjustment costs for capital, with firms investing in a lumpy way and the returns to capital accruing over long periods, changes in capital may not show up immediately in changes in output. Griliches and Mairesse (1998) recommend looking at longer differences, to reduce the role of noisy year-to-year fluctuations. But as noted above (footnote 27), the real-exchange-rate fluctuations that are the main source of exogenous variation in our predicted-import-price instrument have an effect on prices only in the relatively short term, typically 1-2 quarters, and the instrument has little explanatory power over longer periods. If we had an external instrument that generated large changes in capital on a year-by-year basis, it would help greatly, but we have not found such an instrument. In light of these issues, we conclude that we do not have sufficient signal in year-on-year capital changes to estimate β_k well in first-differences. At the same time, we would not feel justified in simply dropping Δk_{it} from the regression (and letting it be incorporated in the error term) because it may still be correlated with changes in materials and labor. Instead, we include Δk_{it} in the first step and treat β_k as a nuisance parameter. In the next subsection, we present a strategy for estimating β_k , in levels, which is arguably less subject to measurement error in the level of capital.

It is worth emphasizing that, under our assumptions, the first step on its own generates consistent estimates of β_m and β_ℓ . If one is only interested in these estimates, for instance to construct markups in the method of Hall (1988) and De Loecker and Warzynski (2012), then one can stop at this step.

2.5.2 Levels (Step 2)

Using $\widehat{\beta}_m$ and $\widehat{\beta}_\ell$ from Step 1, our levels equation, (13), can be rewritten as:

$$\widetilde{y}_{it}^{SV} - \widehat{\beta}_m \widetilde{m}_{it}^{SV} - \widehat{\beta}_\ell \ell_{it} = \beta_k k_{it} + \xi_t + \check{u}_{it} \quad (27)$$

where the error term now includes the firm effect, η_i , and terms arising from estimation error in the first-step coefficients:

$$\check{u}_{it} = \eta_i + (\beta_m - \widehat{\beta}_m) \widetilde{m}_{it}^{SV} + (\beta_\ell - \widehat{\beta}_\ell) \ell_{it} + (\beta_m v_{it}^m - v_{it}^y) + (\beta_m q_{it}^m - q_{it}^y) + \omega_{it} + \epsilon_{it}$$

To justify using a lagged difference of log capital as an instrument for the level of capital, we need an additional assumption. We assume in particular that

$$\mathbb{E}(k_{it}|\eta_i) = c_i \tag{28}$$

for some (potentially firm-specific) constant c_i . This assumption is similar to the standard System-GMM assumption that lagged differences are uncorrelated with the firm effect, which Bun and Sarafidis (2015) note is equivalent to assuming that the correlation of the level of inputs and the firm effect is constant over time. Here we impose the slightly stronger requirement that the conditional expectation of the capital stock is constant, which will be necessary to ensure that the lagged difference is uncorrelated with the levels of the quality and variety terms, q_{it}^m , q_{it}^y , v_{it}^m and v_{it}^y . This assumption rules out correlation between a firm's time-invariant productivity and the evolution of its capital stock over time, although it allows, for instance, current investment to be a function of current and past shocks to productivity.

To proceed, we also need to take a stand on the firm-specific normalizations, \tilde{m}_{i0} and \tilde{y}_{i0} , in (12). This amounts to choosing a base year for the firm-specific output and input price indexes, \tilde{P}_{it} and \tilde{W}_{it}^m . Here we assume that these indexes are equal to unity in the first year that a firm appears in our data. In logs, since $r_{it} = \tilde{y}_{it} + \tilde{p}_{it}$ in every period, setting $\tilde{p}_{i0} = 0$ implies $\tilde{y}_{i0} = r_{i0}$ (where 0 refers to the initial year for the firm). Similarly, on the input side, setting $\tilde{w}_{i0}^m = 0$ implies $\tilde{m}_{i0} = e_{i0}$.

Appendix A.5 shows that, under our assumptions, the lagged difference in capital, Δk_{it-1} , is uncorrelated with \check{u}_{it} and hence is a valid instrument. Intuitively, given that capital can be adjusted only with a lag, and given our assumptions of no serial correlation in the productivity shocks, variety or quality choices conditional on firm fixed effects (see Section 2.4), these terms are redundant in explaining investment choices, conditional on the fixed effect. The constant conditional mean assumption, (28), in turn ensures that Δk_{it-1} is uncorrelated with η_i .

Although the first-step estimation errors in $\hat{\beta}_m$ and $\hat{\beta}_\ell$ show up in \check{u}_{it} , the consistency of the first-step estimates implies that they will not render the second-step estimates inconsistent. But given that there may be correlation between Δk_{it-1} and \tilde{m}_{it}^{SV} or ℓ_{it} , a correction needs to be applied to the standard error for $\hat{\beta}_k$ (Kripfganz and Schwarz, 2019).

If the model is specified correctly, then estimating it in two steps potentially involves a loss of efficiency relative to simultaneous GMM estimation. But as pointed out by Kripfganz and Schwarz (2019), an advantage of the two-step approach is that the first-step estimates of β_m and β_ℓ are robust to mis-specification in the second stage, and in particular to violations of the constant conditional mean assumption, (28).

3 Data, Institutional Background, and Descriptive Statistics

This section describes the main datasets we use, provides institutional background on the minimum wage in Colombia, explains our choice of subsectors, and presents summary statistics for our sample.

3.1 Annual Manufacturing Survey

We use information on sales, employment, wages, capital stock, inputs and outputs from the *Encuesta Anual Manufacturera (EAM, Annual Manufacturing Survey)*, collected by the Colombian statistical agency, known by its Spanish acronym, DANE. Data are reported at the plant level and we aggregate them to the firm level — the level at which we observe imports and exports from trade transactions records (see below). In the sectors we focus on, nearly all firms have just one plant. We focus on data from the period 1994-2009.³⁰ Given that we need at least two lags in our baseline specifications, our main period of analysis is 1996-2009.

The survey contains information on the values and physical quantities of all outputs produced and inputs consumed by each plant at the level of 7-digit Central Product Classification (CPC) categories.³¹ Because the survey is used to construct producer price indexes, DANE pays careful attention to the units of measurement for each product, and a given product is always reported using the same units. We calculate product prices at the firm level as unit values: $P_{ijt} = R_{ijt}/Y_{ijt}$, where R_{ijt} is the value of product k produced by firm i in year t and Y_{ijt} the corresponding quantity. Input prices are calculated analogously. Further details, including on the construction of capital stock, which uses a standard perpetual-inventory method, are in Appendix B.1. The fact that the survey contains, in principle, information on all material inputs is important because it responds to a criticism of IV methods, for instance by Akerberg et al. (2015), that the exclusion restrictions for input-price instruments are likely to be violated if one observes only a subset of inputs.

3.2 Customs Records and Exchange Rates

The customs data contain information from the administrative records filled out by every Colombian importer or exporter for each international transaction, collected by the Colombian customs agency, known by its Spanish acronym, DIAN. Information is available at the level of the firm, product code (8 digit), year, and country of origin (for imports) or destination (for exports). For the period of the analysis, imports and exports by the firm are merged with the EAM manufacturing data using firm identifiers according to the procedures established by DANE. Further details are in Appendix B.2.

³⁰Prior to 1994, the EAM used different plant identifiers and it is often difficult to track plants over time. Although we use data from 1992-1993 when available in constructing firm-level capital stock, we do not focus on these years in the main analysis. Using procedures established by DANE to protect the confidentiality of the data, it is possible to link the customs data (see below) to the EAM only until 2009.

³¹The survey also reports information on outputs sold and inputs purchased, but throughout the paper we use the information on production and consumption to avoid timing issues that arise because firms hold inventories.

To calculate real exchange rates (RERs) by trading partner, we use nominal exchange rates and consumer price indexes (CPIs) from the International Financial Statistics (IFS) of the International Monetary Fund.³² Several of Colombia's most important import origins had significant RER fluctuations over our study period. Venezuela and Mexico, both major oil producers, had large real appreciations in 1995-2000 and large real depreciations subsequently. Indonesia suffered a major crisis accompanied by sharp real devaluation in 1997, as did Argentina in 2001. Even the US and Eurozone countries, which were less volatile overall, experienced non-trivial variation in the RER relative to Colombia.

3.3 Minimum Wage

Despite wide variation in local labor market conditions, Colombia has a single national minimum wage. Over our study period, it was one of the highest in Latin America as a share of the median wage, and it increased significantly in real terms (Maloney and Nuñez Mendez, 2004; Mondragón-Vélez et al., 2010). As required by the Colombian constitution, increases for the coming year were negotiated in December by a tripartite commission including representatives from government, employer associations, and labor organizations. Prior to 1999, the target was commonly understood to be predicted inflation plus predicted productivity growth (Maloney and Nuñez Mendez, 2004). In 1999, because of a recession, predicted inflation greatly exceeded actual inflation and the real value of the minimum wage increased by 7%. In addition, the Constitutional Court in Colombia ruled in 1999 that the minimum wage increase could not be lower than the previous year's inflation. As a result, the real value of the minimum wage continued to rise after 2000, remaining above 90% of the median wage through the end of our study period (Mondragón-Vélez et al., 2010). The real minimum wage increased steadily over 1996-2009, by approximately 23% overall during the period. Appendix Figure A1 plots a histogram of real wages in 1998 for individuals who report working in firms with 10 or more employees in manufacturing in a Colombian household survey, the *Encuesta Nacional de Hogares*. The solid and dashed vertical lines represent the 1998 and 1999 minimum wages, respectively. We see that there was extensive bunching of wages at the minimum in 1998, and that a large share of manufacturing workers was directly affected by the 1999 minimum wage increase. Over the period of analysis, the minimum wage was often used to index the wages of employees who earned above the minimum; as a consequence, increases in the real minimum wage were likely to have an effect on wages throughout the distribution (Mondragón-Vélez et al., 2010). Researchers have previously found disemployment effects of the minimum wage in Colombia, in contrast to several other countries in the region (Bell, 1997; Maloney and Nuñez Mendez, 2004).

³²For a few countries with no information in the IFS, we gathered data directly from their central banks.

3.4 Choice of Subsectors and Descriptive Statistics

Our method is most applicable in industries that meet several criteria. First, the ability to accommodate endogenous quality and variety choices is most valuable in sectors producing differentiated products, particularly those with substantial quality variation. Second, given that we assume that firms are price-takers in input markets, our method is most applicable in industries in which inputs, although they may differ in quality, are relatively non-specialized and substitutable within quality categories. Third, for our external instrument for materials to be relevant, a substantial share of inputs in the industry must be imported, such that real-exchange-rate fluctuations have a significant effect on the input prices faced by firms.

In choosing subsectors that fit these criteria, we face a familiar trade-off. On one hand, we would like sample sizes to be as large as possible in order to increase the precision of our estimates. This clearly matters in our setting where the weakness of instruments is a concern. On the other hand, the wider the net that we cast, the more heterogeneous the included firms are likely to be. The issue is particularly salient because, as is common in the literature, we will treat all firms in our sample as having the same production-function coefficients.

Our approach in this paper is to focus on producers of rubber and plastic products. These subsectors are adjacent in the ISIC revision 2 classification (with 3-digit codes 355 and 356, respectively) and are often classified together in a 2-digit sector, as for instance in Sector 36 (“Rubber and Plastic Products”) of the U.N. Central Product Classification (CPC). Table A1 reports their main 8-digit outputs. For rubber, the main product is tires of different kinds. These can be understood to be differentiated products: they are typically sold under brand names — Goodyear and Michelin tires are produced in Colombia, for instance — and often for fairly specialized uses. For plastics, there is less concentration on a single type of product; output is distributed across various types of tubing, bags, sheets, films, and containers. But again, the products are typically differentiated and often tailored for specialized uses.

By contrast, the inputs of both subsectors can be viewed as commodities, or at least commodity-like — highly substitutable across suppliers even if they have quality differences. Table A2 reports the main 8-digit inputs. For rubber, the most important input is natural latex, from the bark of rubber trees. The second-most important input category, “rare metals in primary forms” (CPC product code 3423112), includes carbon black, a form of carbon used as a filler in tires. For the plastics subsector, the most important inputs are raw forms of different common plastics — polyethylene (PET), polypropylene (PP), polyvinyl chloride (PVC), polystyrene, and others — often purchased in the form of pellets. Although pellets vary in their chemical properties, their characteristics are typically noted on the packaging. Within a given chemical specification, pellets from different producers and origin countries are typically considered to be highly substitutable. There may be other dimensions of supply relationships that cannot be observed *ex ante*, for instance timeliness of deliv-

ery or willingness of supplier to extend trade credit. But to a first approximation we believe it is reasonable to treat the main inputs in rubber and plastics as highly substitutable, with observable quality differences.

As is evident in Table A2, a large share of inputs in both subsectors is imported. In rubber products, almost all natural latex is imported, as are substantial shares of carbon black and other inputs. In plastics, a majority of PET and 20-25% of PVC and polystyrene are imported. These import shares are from the EAM data and hence represent shares of inputs imported directly by firms. To the extent that firms purchase imported goods from local intermediaries, they understate the true import shares of the inputs.

In selecting the estimation sample, we require that a firm have complete data on capital, labor, materials, and outputs for at least six consecutive years. This requirement is helpful to ensure that the perpetual-inventory method generates a sensible measure of capital stock. It also ensures that our sample of firms does not change as we modify the number of lags required in different specifications. We are left with 362 firms in an unbalanced panel, with 11.7 observations per firm on average over 1996-2009. Table A3 presents summary statistics on this baseline sample. We see that the two subsectors are comparable on many dimensions.

To explore the robustness of our results, in the appendix we will report results for two alternative samples. In the first, we remove rubber products and focus exclusively on plastics, the larger of the two subsectors. In the second, we keep both rubber and plastics and add glass products (ISIC rev. 2 subsector 362), a subsector that also arguably satisfies the criteria of relatively substitutable inputs, high imported input share, and differentiated outputs. Descriptive statistics on the glass products subsector are reported in supplementary materials available from the authors.

4 Results

This section reports the results of the estimation strategy laid out in Section 2. For comparison purposes, we begin by presenting “naive” OLS and first-difference (FD) results, and then move on to our two-step IV (TSIV) method.

4.1 “Naive” OLS and FD Estimators

Panel A of Table 1 presents estimates using (deflated) sales as the measure of output and (deflated) material expenditures as the measure of input use. The OLS estimates in Columns 1 and 2, without and with year effects respectively, appear to be reasonable, and are roughly consistent with constant returns to scale, as is typically expected (see e.g. Bartelsman and Doms (2000)). Columns 3 and 4 report first-difference (FD) estimates, corresponding to equation (14) without instruments. Relative to the OLS estimates, the materials coefficients are significantly lower, the labor coefficients remain roughly unchanged, and, strikingly, the capital coefficients drop almost to zero. The latter fact

is consistent with the observation that transformations to remove firm effects can lead to severe attenuation of the capital coefficient; this problem is not specific to our TSIV method.

Panel B of Table 1 again reports OLS and FD estimates, but using the Sato-Vartia quantity aggregates for output and materials. In Columns 1 and 2, we have imposed the firm-specific normalizations for \tilde{y}_{i0} and \tilde{m}_{i0} discussed in Section 2.5.2 above, using each firm’s first year in the unbalanced panel as the base year for the firm-specific output and input deflators.³³ Comparing Panels A and B, we see significant differences in the OLS estimates — in particular, using the quantity indexes reduces the OLS materials coefficient and raises the capital coefficient — but the FD estimates are quite similar across panels.

4.2 Differences (Step 1) Results

In this step, we estimate our differences equation, (14), using instruments for the changes in input choices. Table 2 reports the first stage for different sets of instruments. Columns 1-3 use only internal instruments, and in particular only lagged levels of inputs from period $t-2$. The coefficient estimates are plausible, with lagged levels negatively associated with current changes.

Testing for weak instruments is complicated in this setting by the presence of multiple endogenous covariates and the potential for heteroskedastic errors (which we would not feel justified in assuming away). This is a frontier area of econometric theory and there is no consensus on the right diagnostic tests for such cases.³⁴ Two tests are commonly reported in practice. Sanderson and Windmeijer (2016) propose an improved version of a test first suggested by Angrist and Pischke (2009), which is appropriate for inference on each of multiple endogenous regressors.³⁵ Also commonly reported is the Kleibergen and Paap (2006) Wald statistic, an analogue of the Cragg and Donald (1993) statistic applicable in non-homoskedastic settings. Using these diagnostics, there is evidence that the instruments are weak. The Sanderson-Windmeijer (SW) F-statistics are well below the rule-of-thumb level of 10 (as are the conventional F-statistics for materials and labor), and although the Kleibergen-Paap (KP) LM test easily rejects the null of underidentification, the KP Wald statistic for weak instruments is below 1.³⁶ In Appendix Table A4, we show that this weak-instrument issue is not resolved by including further lags as instruments in a GMM estimator.³⁷

³³These normalizations are differenced out in the first-difference specifications.

³⁴In a recent state-of-the-art review, Andrews et al. (2019) recommend the test of Montiel Olea and Pflueger (2013) in cases with a single endogenous regressor, but have no recommendation in cases with multiple endogenous regressors; see their footnote 4.

³⁵The Sanderson-Windmeijer statistic adjusts for the fact that the endogenous covariates may themselves be highly correlated. The theoretical justification for it relies on an assumption of homoskedastic errors, but it is commonly reported even in non-homoskedastic settings.

³⁶Although the Kleibergen and Paap (2006) Wald statistic is sometimes compared to the Stock and Yogo (2005) critical values, Andrews et al. (2019) note that this comparison lacks theoretical justification in the non-homoskedastic case. We simply note that this statistic is at a level that would typically raise concerns among practitioners.

³⁷Appendix Table A4 reports results from GMM estimation of our difference equation (14), where further lags have been added “GMM-style” (Holtz-Eakin et al., 1988; Roodman, 2009), allowing separate coefficients in each period. Lags are included just from $t-2$ in Column 1, from $t-2$ and $t-3$ in Column 2, and from $t-2$ to the firm’s initial

To improve the explanatory power of the first stage of this step, we turn to our external instruments. As described in Section 2.5.1 above, we start by estimating the relationship between RER movements and import prices given by (23), leaving out one firm at a time. This generates 362 sets of coefficient estimates, one for each firm in our sample. (The results are reported in supplementary materials available from the authors.) Although there is some heterogeneity, in the majority of sectors, and on average across sectors, import prices are positively related to RER movements, as expected.

Columns 4-6 of Table 2 report the first-stage estimates including the two external instruments — the predicted change in import price, $\Delta \widehat{w}_{int}^{imp}$ from (24), and the minimum wage instrument, Δz_{it} from (26) — and one internal instrument, the lagged level of capital from $t - 2$. The coefficient estimates broadly conform to our expectations. In particular, the predicted import price change is significantly negatively related to the change in the material quantity aggregate and the predicted wage change is significantly negatively related to the change in employment. In the latter case, the predicted wage change is also negatively related to the materials and capital changes. The instruments are somewhat stronger than in the internal-instruments-only model in Columns 1-3, but both the SW F-statistic for materials and labor and the KP Wald statistic continue to warrant concern about the weakness of the instruments.

Our preferred specification combines the three internal instruments from $t - 2$ and the external instruments. The corresponding first stage is reported in Columns 7-9 of Table 2. The coefficient estimates are similar to those in the other columns but the strength of the instrument set has improved. The SW F-statistic is above the rule-of-thumb level of 10 for labor and capital and the KP Wald statistic, while still below 3, is larger than in the other columns. The concern about the weakness of instruments remains, but it has been mitigated by the inclusion of the external instruments.

Table 3 presents the second-stage estimates corresponding to the three sets of instrument in Table 2. In the first two columns, the coefficients on materials and labor are imprecisely estimated and change markedly across columns, as one might expect given the weakness of the instruments in these specifications. In our preferred specification in Column 3, by contrast, the materials and labor coefficients are more precisely estimated and are of plausible magnitudes, 0.45 and 0.47 respectively. The labor coefficient is substantially larger than, and the materials coefficient very similar to, the corresponding FD estimates in Table 1, Panel B, Columns 3-4. The difference in the labor coefficient is consistent with the presence of an output-quality bias discussed above: if producing higher-quality output requires more labor, then we would expect a positive correlation between $\Delta \ell_{it}$ and Δq_{it}^y in (14), generating a negative bias in OLS and FD estimates of β_ℓ , which our approach would correct.

As previewed above, the capital coefficient is implausibly low in this specification. The point year in the sample in Column 3. The Kleibergen-Paap Wald statistic remains below 2 and the Sanderson-Windmeijer F statistics are all below 3.5.

estimate is in fact negative, although the confidence interval allows for positive values of roughly the magnitude of the OLS estimate in Columns 3-4 of Table 1, Panel A. In Step 2 below, using the levels equation, we will arrive at a more plausible point estimate for the capital coefficient.

Given that the SW F-statistic for materials and the KP Wald statistic are still somewhat low in our preferred specification, we explore the robustness of the estimates in two ways. First, we report weak-instrument-robust confidence intervals. The econometric literature has not reached consensus on the best method for estimating these intervals, especially in the non-homoskedastic case. Here we follow the approach of Andrews (2016, 2018), which uses a statistic based on a linear combination (LC) of the K statistic of Kleibergen (2005) and the S statistic of Stock and Wright (2000). We treat β_k as a nuisance parameter and do not assume that it is strongly identified. The confidence intervals for β_m and β_ℓ are reported in Column 3 of Table 2. The intervals are centered at the reported point estimates and allow us to reject the nulls that $\beta_m = 0$ and $\beta_\ell = 0$ comfortably at the 95% level. Second, to further probe robustness, we estimate the Column 3 specification using limited-information maximum likelihood (LIML), which has been found to be more robust to weak instruments than IV (Stock et al., 2002; Angrist and Pischke, 2009). The Andrews LC robust confidence intervals, reported in Column 4, are somewhat larger, but the coefficient estimates are nearly identical to those in Column 3, which is reassuring.

Table 3 also reports the Arellano and Bond (1991) test statistics for serial correlation in the residuals of the difference equation. We easily reject that there is no correlation between Δu_{it} and Δu_{it-1} , unsurprisingly, but we fail to reject the null of no correlation between Δu_{it} and Δu_{it-2} . As discussed in Section 2.4 above, this is consistent with the assumptions required for our internal instruments to be valid in this context.

4.3 Levels (Step 2) Results

We now turn to the second step of our TSIV procedure, in levels. We estimate equation (27), where we have plugged in $\widehat{\beta}_m$ and $\widehat{\beta}_\ell$ from the first step on the left-hand side.

Panel A of Table 4 reports the first stage of the IV procedure for this step, using Δk_{it-1} as the instrument for k_{it} . Weakness of the instrument is not a concern here: the Kleibergen-Papp Wald statistic is above 39.³⁸ Although the R-squared is low, the first-stage coefficient is 0.67 and highly significant.

Panel B of Table 4 reports the second stage in Column 1 and, for comparison purposes, the corresponding OLS estimate in Column 2. The square brackets in Column 1 report the corrected standard error discussed in Section 2.5.2 above. The corrected standard error does not allow us to reject the null that $\beta_k = 0$ at conventional levels, but the point estimate for the capital coefficient of 0.11 is plausible and, together with the first-step estimates of $\widehat{\beta}_m$ and $\widehat{\beta}_\ell$, 0.45 and 0.47, indicates

³⁸The Kleibergen-Papp Wald statistic and the Sanderson-Windmeijer F statistic coincide in cases with a single endogenous covariate.

that returns to scale are nearly constant, as generally expected (Bartelsman and Doms, 2000). While one would prefer to have a more precise estimate of β_k , we have more confidence in the estimate than in the close-to-zero estimates from first-differences (Columns 3-4 of Panels A and B of Table 1) or the negative estimates from the first step of our TSIV procedure (Table 3).

4.4 Robustness

Appendix C reports on three exercises to probe the robustness of our estimates: using alternative aggregators (Tornqvist, Paasche, and Laspeyres indexes), adding a control for the firm-level predicted export price as a covariate, and using alternative samples (dropping rubber product producers or adding glass product producers). The estimates reported above are largely robust to these changes.

5 Comparison to Other Methods

This section compares our output-elasticity estimates to those of other commonly used methods. We begin by comparing estimates across methods in the Colombian data (Subection 5.1) and then explore the performance of the various methods in a simple Monte Carlo simulation (Subsection 5.2).

5.1 In Colombian Data

To compare our estimates to those of System GMM, we implement System GMM using our Sato-Vartia quantity indexes. (Since the System GMM set-up includes a firm fixed effect, it can absorb the firm-specific normalizations we require for the quantity aggregates.) Table 5 presents the results. We include time fixed effects and use the “two-step” procedure described in Roodman (2009).³⁹ The coefficients on the contemporaneous materials quantity index (which is logged), log labor, and log capital are estimates of the Cobb-Douglas output elasticities, corresponding to our β_m , β_ℓ , and β_k .⁴⁰ The columns differ in the number of covariate lags included in the difference equation, with lags just from $t - 3$ in Column 1, from $t - 3$ and $t - 4$ in Column 2, and from all available periods $t - 3$ and earlier in Column 3. The instruments are included “GMM-style,” effectively interacted with year dummies (Holtz-Eakin et al., 1988; Roodman, 2009). In the corresponding levels equations, we include the first lags of the first-differenced covariates as instruments.⁴¹ To gauge the strength of the System GMM instruments, we follow Bazzi and Clemens (2013) and Kraay (2015) in reporting weak-instrument diagnostics separately for the differences and levels equations in Appendix Table

³⁹In particular, we use the Stata `xtabond2` command of Roodman (2009) with options `h(2)`, `twostep`, and `robust`. Following Roodman’s replication of Blundell and Bond (1998), we include time fixed effects as instruments only in the levels equation, since they are asymptotically redundant in the difference equation.

⁴⁰The model implies additional restrictions on the relationship between the coefficients on the contemporaneous and lagged terms, which we do not test here.

⁴¹Further lags of differences in the levels equation are redundant (Arellano and Bover, 1995; Blundell and Bond, 2000).

A14. The diagnostics give some reason for concern about the strength of the instruments, with the Sanderson-Windmeijer F-statistics below 2.3 for contemporaneous materials, labor and capital.⁴² The materials (0.455) and capital (0.106) coefficients from the “all lags” specification (Column 3) are quite similar to our TSIV estimates (0.450 and 0.114, respectively). The main difference arises in the labor coefficients, where the System GMM coefficient (0.292) is smaller than ours (0.472). Given the large weak-instrument-robust confidence interval around our estimate, this difference is not statistically significant, but it does suggest that the inclusion of the external instruments to strengthen the instrument set may be important in this case.

For the other methods we consider, we rely on (deflated) revenues and expenditures, the variables that are available in standard datasets. Table 6 reports estimates from the methods of Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), Wooldridge (Wooldridge, 2009), Gandhi, Navarro and Rivers (2020) (GNR), and an extension of GNR that allows for monopolistic competition in output markets, which we label GNR-MC.⁴³ The method of Akerberg, Caves and Frazer (2015) (ACF) is also commonly used, but the authors recommend that it only be used with value-added production functions, not gross output functions, hence coefficient estimates from their method would not be directly comparable to ours and we omit them here. We include our TSIV estimates in Column 6 for comparison purposes. With the caveat that our confidence intervals are wide, some differences in the point estimates are worth noting. For OP and LP in Columns 1-2, the point estimates for materials are consistently higher, and for labor are consistently lower, than our estimates. The GNR estimates in Column 3 are closer to ours. The GNR-MC estimates, which scale up the coefficients using a markup estimate, are larger than ours for all three inputs. It is worth noting that while the differences in point estimates are generally not statistically significant, the point estimates will matter, potentially in important ways, for the estimates of markups and productivity that are constructed using them.

5.2 Monte Carlo Simulation

While the comparison in the previous subsection reveals several interesting patterns, the interpretation is made difficult by the fact that we do not know the true values of the output elasticities. As a complementary exercise, here we present a simple Monte Carlo simulation, comparing the estimators in an artificial setting we understand well. The data-generating processes (DGPs) we consider are consistent with the assumptions we have set out above, and it is perhaps not surprising that our es-

⁴²The Hansen test of over-identifying restrictions is appropriate in the non-homoskedastic case and does not reject the hypothesis that the instruments are jointly valid. But it should be interpreted with caution, as it is weakened by the presence of many instruments (Roodman, 2009).

⁴³For OP, LP and Wooldridge, we use the Stata command `prodest` (Rovigatti and Mollisi, 2018) and include year effects; for the GNR estimators, we have coded the estimation ourselves. In implementing GNR, given the Cobb-Douglas structure of the the production function, we use a polynomial of degree zero for the materials expenditure elasticity, a polynomial of degree one in capital and labor for the integration constant, and a polynomial of degree three for the AR(1) process of ω_{it} . For all specifications we obtain standard errors by using a bootstrap with 50 replications.

timator performs well under them. But we nonetheless believe it is instructive to consider the relative performance of different estimators in the presence of features that are ruled out by the assumptions required for other common methods. The details of the DGPs and additional results are presented in Appendix D. Here we summarize the main patterns.

To make the simulation as simple and transparent as possible, we impose a number of restrictions on our theoretical framework. We assume that firms use a single material input, along with capital, to produce a single output of homogeneous quality, abstracting from variety effects, output quality effects, and labor choices. We focus on the role of input quality, allowing it to depend on exogenous shocks as well as endogenous choices of firms (based on their ex-ante productivity shocks). To the extent possible, we follow the simulation in Gandhi et al. (2020). In our setting we must allow for variation in input prices over time in order for our internal instruments to have explanatory power for contemporaneous choices. We allow materials prices to have an international component (used as an external instrument in our TSIV procedure) and a domestic component. We consider four DGPs. We begin with a simple DGP with perfect competition and serially uncorrelated productivity shocks, which we label DGP1. We then change to monopolistic competition (in DGP2), add time-invariant firm effects (DGP3), and add input-quality variation (in DGP4). For each DGP, we construct 100 samples of 500 firms over 30 periods. The true values of the materials and capital coefficients are 0.65 and 0.25, respectively. We assume that output and input quantities are observable and use them throughout (except for the first step of GNR and for GNR-MC as explained in Appendix D). Note that the presence of monopolistic competition may still affect input and output quantities.

Table 7 presents the results, reporting estimates of the methods considered above. In parentheses we report the standard deviations of the coefficient estimates across the 100 samples for each DGP; these are effectively bootstrap estimates of the standard errors of the coefficients. OLS rejects the true values at the 95% level for all DGPs. For DGP1 and DGP2, the issue is standard transmission bias.⁴⁴ The inclusion of firm effects in DGP3 increases the cross-sectional variance of known-ex-ante productivity, and hence exacerbates the transmission bias in this case. Input-quality bias in DGP4 shifts the estimates further from their true values; this is expected because higher input quality both leads firms to use more materials and contributes to higher output. The FD estimates are similar to OLS, but (naturally) are less affected by the inclusion of fixed effects. On the other hand, the first-differencing removes signal in materials and capital, and the coefficients are less precisely estimated. The System GMM point estimates are similar to OLS, but the standard errors are (appropriately) larger and the 95% confidence intervals include the true values. Given the relatively low degree of serial correlation we have built into the simulation, it appears that the internal instruments that System GMM relies on are not on their own strong enough to generate precise estimates.

OP performs reasonably well overall: it does not reject the true values at conventional levels

⁴⁴Because the firm’s optimal materials choice in DGP2 depends on the intensity of competition (captured by σ), the extent of transmission bias is also sensitive to σ .

of confidence for any of the DGPs and yields more precise estimates than System GMM. Although the firm effects in principle violate the monotonicity property required to construct the proxy for productivity, in this simulation the variance of the firm effects is sufficiently low that the consequences are not severe. However, two words of caution are in order. First, the introduction of input-quality differences increases the materials coefficient, as we would expect given that we have assumed that firm optimally consume more of a higher-quality input (holding input price constant), generating a positive input-quality bias. Greater variance in input quality would exacerbate this bias. Second, the OP method has the shortcoming, noted by Levinsohn and Petrin (2003), that it requires dropping observations with zero investment. Given we have not built any heterogeneity in output elasticities into the simulation, this sample selection does not greatly affect the coefficient estimates, but it should remain an important consideration in other contexts.

Similarly to the other methods discussed so far, the LP estimates move further from the true values when moving from DGP2 to DGP3 to DGP4.⁴⁵ The introduction of input-quality differences has a more direct effect on materials choices than investment, and hence here has a bigger impact on estimates using materials as a proxy for productivity (LP) than those using investment (OP).

Turning to the GNR and GNR-MC estimates in Columns 6 and 7, we see that both methods precisely recover the true elasticities for DGP1. The difficulties for these methods arise when monopolistic competition is introduced. Gandhi et al. (2020) themselves show that under monopolistic competition the standard GNR estimate of the materials coefficient converges to the true elasticity times one over the true markup, which in our case is set to 1.18 (see Appendix D). The GNR-MC extension seeks to correct the estimates using an estimate of the markup. Following De Loecker (2011), it uses a market-share-weighted average of deflated revenues as an aggregate demand shifter to construct this estimate. In our context, variation in the aggregate demand shifter comes from variation in household income. The Column 7 Panels B-D estimates indicate that this correction for the markup is not sufficient in our setting. Greater variation in aggregate household income would improve the markup estimate, but in real data it will be difficult to know whether the GNR-MC correction is adequate in imperfectly competitive settings. Although in principle the standard versions GNR and GNR-MC are not able to accommodate firm fixed effects,⁴⁶ in this simulation the variance in the fixed effects is not large enough to materially affect the estimates. It is noteworthy that the GNR and GNR-MC estimates are not adversely affected by the addition of input-quality differences.⁴⁷

Column 8 reports our TSIV estimates, using the second lags of materials and capital as internal

⁴⁵Although we do not have not presented results using the ACF method because we have been focused on gross output functions, we note that in this simulation, which omits labor, if ACF were applied it would coincide with LP.

⁴⁶See Section VII.C.3 of Gandhi et al. (2020) for a discussion of an extension to accommodate firm fixed effects.

⁴⁷In this context, the input quality shocks can be interpreted as simply adding to the variance of productivity shocks (as long as the sum of the two shocks is itself a Markov process). Since GNR do not attempt to proxy for productivity directly with investment or materials demand, their method is robust to such changes.

instruments, and the log change of the international component of material’s price as an external instrument. (See also the discussion in Appendix D.2.) Our estimates are robust to the inclusion of imperfect competition, firm effects, and input-quality differences. Overall, our method performs better than the other methods we consider. While the simulation environment is admittedly artificial, it does point to advantages of our method in differentiated-product sectors with input-quality differences. Although in the interests of simplicity we have not included variety differences or output-quality differences in the simulation, it seems clear that they would further accentuate the advantages of our approach.

6 Conclusion

This paper has sought to make three main contributions. First, we have highlighted the pitfalls of using physical quantities to estimate output elasticities in industries where quality and variety vary across firms and over time. Using within-firm CES aggregators for outputs and material inputs, we have shown theoretically how standard estimates of output elasticities are likely to be biased by quality and variety differences. Second, we have used external drivers of input prices as instruments, in addition to standard internal instruments, to estimate the elasticities of output with respect to materials and labor. The idea of using external instruments is not new, but previous authors have not had access to the combination of datasets and naturally occurring variation that we have been able to take advantage of here. Third, we have developed a new approach to estimating of the output elasticity with respect to capital. It is well known that first-difference estimators tend to yield unsatisfactory estimates of the capital coefficient, and our first-step (differences) estimates are no exception. In the spirit of System GMM, we have added a levels equation, using a lagged difference as an instrument. Our two-step IV (TSIV) approach has the advantage that the estimates of the materials and labor coefficients are robust to misspecification of the second step, which requires stronger assumptions. A simple Monte Carlo exercise suggests that our estimator performs better than standard estimators at least in the artificial environment of the simulation.

An important question that remains open is what to do if one does not have the combination of rich data and naturally occurring variation that we have in our setting. Data on physical quantities of both inputs and outputs are increasingly available, in several countries. But it may be difficult to find credible external instruments and internal instruments may be weak. One potential way forward would be to construct proxies for the quality and variety terms that appear in estimating equations such as (13) above. The approach of De Loecker et al. (2016), of including a flexible function of output price and market share on the right-hand-side, is a promising step in this direction. One could also consider constructing explicit measures of quality, as Khandelwal et al. (2013), Piveteau and Smagghue (2019) and Errico and Lashkari (2022) do for output quality; such a strategy would require imposing more structure on consumer demand than we have been willing to do here, but

may be warranted in some circumstances. To proxy for variety, one could include the observable components of the variety terms derived above and allow for firm-specific coefficients on them.⁴⁸ Another potential way forward would be to intensify the search for credible external instruments. We have taken first step in this direction, and we hope we have made some progress in convincing the reader of the desirability of such instruments, but surely more and better external instruments will be found as richer and richer micro-data become available.

Our main objective in this paper has been to improve estimates of output elasticities. We leave for future work the task of putting the estimates to use, in particular to construct improved estimates of markups and productivity. For markups, estimates from our method can be inserted directly into calculations based on Hall (1988) and De Loecker and Warzynski (2012). For this purpose, our estimates have the advantage that they reflect elasticities of real output, not sales or value-added, which Bond et al. (2021) have recently argued is crucial. For productivity, the task is somewhat less straightforward. Two types of TFP measures have received particular attention: TFPQ (Q for quantities), which uses physical output and input quantities, and TFPR (R for revenues), which uses revenues for output and expenditures for inputs. In Appendix E, we show that neither captures only technical efficiency, even when calculated using our improved elasticity estimates. In particular, while TFPQ avoids input- and output-price biases, it may still reflect quality or variety differences on the input or output side. By contrast, the use of revenues and expenditures in TFPR avoids what we have called quality and variety biases, but it may reflect idiosyncratic variation in output or input prices at the firm level. In settings where quality or variety differences are important (such as ours in this paper), we would argue that TFPR should be preferred, keeping the caveat from the previous sentence squarely in mind. Appendix Table A15 shows that TFPR constructed using our output-elasticity estimates is correlated with TFPR calculated using the other methods considered in Section 5 above, but the correlation is far from perfect. The choice of estimator is likely to matter greatly for calculations such as the effect of trade on productivity or the extent of productivity dispersion, the basis for recent estimates of misallocation in Hsieh and Klenow (2009) and related literature. Our method may be useful for a variety of applications that require estimates of markups or productivity.

References

- Abito, Jose Miguel**, “Estimating Production Functions with Unobserved Time-Invariant Heterogeneity and Time-Varying Productivity Shocks,” 2021. Unpub. paper, University of Pennsylvania.
- Ackerberg, Daniel A**, “Comment on “Olley and Pakes-style Production Function Estimators with Firm Fixed Effects”,” *Oxford Bulletin of Economics and Statistics*, 2021, 83 (3), 836–840.

⁴⁸Note that the variety terms in (13) depend only on observables (the χ terms) and time-invariant constants (the elasticities of substitution, σ_i^y and σ_i^m).

- Ackerberg, Daniel A., Kevin Caves, and Garth Frazer**, “Identification Properties of Recent Product Function Estimators,” *Econometrica*, 2015, 83 (6), 2411–2451.
- Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes**, “Econometric Tools for Analyzing Market Outcomes,” in James Heckman and Edward Leamer, eds., *Handbook of Econometrics*, vol. 6A, Amsterdam: Elsevier, 2007, chapter 63, pp. 4172–4276.
- Amiti, Mary, Oleg Itskhoki, and Jozef Konings**, “International Shocks, Variable Markups, and Domestic Prices,” *The Review of Economic Studies*, 2019, 86 (6), 2356–2402.
- Andrews, Isaiah**, “Conditional Linear Combination Tests for Weakly Identified Models,” *Econometrica*, 2016, 84 (6), 2155–2182.
- , “Valid Two-Step Identification-Robust Confidence Sets for GMM,” *Review of Economics and Statistics*, 2018, 100 (2), 337–348.
- , **James H Stock, and Liyang Sun**, “Weak Instruments in Instrumental Variables Regression: Theory and Practice,” *Annual Review of Economics*, 2019, 11, 727–753.
- , **Nano Barahona, Matthew Gentzkow, Ashesh Rambachan, and Jesse Shapiro**, “Included and Excluded Instruments in Structural Estimation,” 2022. Unpub. paper, Harvard University.
- Angrist, Joshua D. and Jorn-Steffen Pischke**, *Mostly Harmless Econometrics: An Empiricist’s Companion*, Princeton University Press, 2009.
- Arellano, Manuel and Olympia Bover**, “Another Look at the Instrumental Variable Estimation of Error-Components Models,” *Journal of Econometrics*, 1995, 68 (1), 29–51.
- **and Stephen Bond**, “Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations,” *Review of Economic Studies*, 1991, 58 (2), 277 – 297.
- Atkin, David, Amit K. Khandelwal, and Adam Osman**, “Measuring Productivity: Lessons from Tailored Surveys and Productivity Benchmarking,” *AEA Papers and Proceedings*, 2019, 109, 444–49.
- Bartelsman, Eric J. and Mark Doms**, “Understanding Productivity: Lessons from Longitudinal Micro-data,” *Journal of Economic Literature*, September 2000, 38 (3), 569–594.
- Bastos, Paulo, Joana Silva, and Eric Verhoogen**, “Export Destinations and Input Prices: Evidence from Portugal,” *American Economic Review*, 2018, 108 (2), 353–392.
- Bazzi, Samuel and Michael A Clemens**, “Blunt instruments: Avoiding common pitfalls in identifying the causes of economic growth,” *American Economic Journal: Macroeconomics*, 2013, 5 (2), 152–86.
- Bell, Linda**, “The Impact of Minimum Wages in Mexico and Colombia,” *Journal of Labor Economics*, 1997, 15, s102–s123.
- Bernard, Andrew B., Stephen J. Redding, and Peter K. Schott**, “Multi-Product Firms and Trade Liberalization,” *Quarterly Journal of Economics*, 2011, 126 (3), 1271–1318.
- Berry, Steven, Amit Gandhi, and Philip Haile**, “Connected Substitutes and Invertibility of Demand,” *Econometrica*, 2013, 81 (5), 2087–2111.
- Blum, Bernardo S, Sebastian Claro, Ignatius Horstmann, and David A Rivers**, “The ABCs of Firm Heterogeneity: The Effects of Demand and Cost Differences on Exporting,” *Journal of Political Economy*, forthcoming.
- Blundell, Richard and Stephen Bond**, “Initial Conditions and Moment Restrictions in Dynamic Panel Data Models,” *Journal of Econometrics*, 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, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch**, “Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data,” *Journal of Monetary Economics*, 2021, 121, 1–14.
- Bun, Maurice J. G. and Frank Windmeijer**, “The Weak Instrument Problem of the System GMM Estimator in Dynamic Panel Data Models,” *Econometrics Journal*, 2010, 13 (1), 95–126.
- **and Vasilis Sarafidis**, “Dynamic Panel Data Models,” *Oxford Handbook of Panel Data*, 2015, pp. 76–110.
- Campa, Jose Manuel and Linda S. Goldberg**, “Exchange Rate Pass-Through into Import Prices,” *Review of Economics and Statistics*, 2005, 87 (4), 679–690.
- Card, David**, “Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage,” *Industrial and Labor Relations Review*, 1992, 46 (1), pp. 22–37.
- Cengiz, Doruk, Arindrajit Dube, Attila Lindner, and Ben Zipperer**, “The Effect of Minimum Wages on Low-Wage Jobs*,” *Quarterly Journal of Economics*, 05 2019, 134 (3), 1405–1454.

- Chamberlain, Gary**, “Multivariate Regression Models for Panel Data,” *Journal of Econometrics*, 1982, 18, 5–46.
- Collard-Wexler, Allan and Jan De Loecker**, “Production Function Estimation and Capital Measurement Error,” 2020. Unpub. paper, Duke University.
- Cragg, John G. and Stephen G. Donald**, “Testing Identifiability and Specification in Instrumental Variable Models,” *Econometric Theory*, 1993, 9 (2), 222–240.
- De Loecker, Jan**, “Product Differentiation, Multi-product Firms and Estimating the Impact of Trade Liberalization on Productivity,” *Econometrica*, 2011, 79 (5).
- and **Chad Syverson**, “An Industrial Organization Perspective on Productivity,” 2021. NBER working paper no. 29229.
- and **Frederic Warzynski**, “Markups and Firm-level Export Status,” *American Economic Review*, 2012, 102 (6), 2437–2471.
- and **Pinelopi Koujianou Goldberg**, “Firm Performance in a Global Market,” *Annual Review of Economics*, 2014, 6 (1), 201–227.
- , **Pinelopi K. Goldberg**, **Amit K. Khandelwal**, and **Nina Pavcnik**, “Prices, Markups and Trade Reform,” *Econometrica*, 2016, 84 (2), 445–510.
- Demirer, Mert**, “Production Function Estimation with Factor-Augmenting Technology: An Application to Markups,” 2022.
- Dhyne, Emmanuel, Amil Petrin, Valerie Smeets, and Frederic Warzynski**, “Theory for Extending Single-Product Production Function Estimation to Multi-Product Settings,” 2022. NBER working paper no. 30784.
- , —, —, and —, “Multi Product Firms, Import Competition, and the Evolution of Firm-Product Technical Efficiencies,” 2023. Unpub. paper, Aarhus University.
- Doornik, Jurgen A, Manuel Arellano, and Stephen Bond**, “Panel Data Estimation Using DPD for Ox,” 2012. Available at https://www.doornik.com/download/oxmetrics7/Ox_Packages/dpd.pdf.
- 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,” *Journal of Political Economy*, 2018, 126 (3), 1027–1084.
- Eckel, Carsten and J. Peter Neary**, “Multi-Product Firms and Flexible Manufacturing in the Global Economy,” *Review of Economic Studies*, Jan 2010, 77, 188–217.
- Errico, Marco and Danial Lashkari**, “The Quality of US Imports and the Consumption Gains from Globalization,” 2022. Unpub. paper, Boston College.
- Eslava, Marcela, John C. Haltiwanger, and Nicolás Urdaneta**, “The Life-Cycle Growth of Plants: The Role of Productivity, Demand and Wedges,” *Review of Economic Studies*, forthcoming.
- , **John Haltiwanger, Adriana Kugler, and Maurice Kugler**, “The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia,” *Journal of Development Economics*, 2004, 75, 333–371.
- , —, —, and —, “Trade Reforms and Market Selection: Evidence from Manufacturing Plants in Colombia,” *Review of Economic Dynamics*, 2013, 16, 135–158.
- Feenstra, Robert C.**, “New Product Varieties and the Measurement of International Prices,” *American Economic Review*, 1994, 84 (1), 157 – 177.
- Forlani, Emanuele, Ralf Martin, Giordiano Mion, and Mirabelle Muûls**, “Unraveling Firms: Demand, Productivity and Markups Heterogeneity,” *Economic Journal*, forthcoming.
- Foster, Lucia, Cheryl Grim, John Haltiwanger, and Zoltan Wolf**, “Firm-level Dispersion in Productivity: Is the Devil in the Details?,” *American Economic Review*, 2016, 106 (5), 95–98.
- , **John Haltiwanger, and Chad Syverson**, “Reallocation, Firm Turnover and Efficiency: Selection on Productivity or Profitability?,” *American Economic Review*, 2008, 98 (1), 394–425.
- Fox, Jeremy T. and Valérie Smeets**, “Does Input Quality Drive Measured Differences in Firm Productivity?,” *International Economic Review*, 2011, 52 (4), 961–989.
- Gandhi, Amit, Salvador Navarro, and David Rivers**, “On the Identification of Gross Output Production Functions,” *Journal of Political Economy*, 2020, 128 (8), 2973–3016.
- Garcia-Marin, Alvaro and Nico Voigtländer**, “Exporting and Plant-Level Efficiency Gains: It’s in the Measure,” *Journal of Political Economy*, 2019, 127 (4), 1777–1825.

- Goldberg, Pinelopi K., Amit Khandelwal, Nina Pavcnik, and Petia Topalova**, “Imported Intermediate Inputs and Domestic Product Growth: Evidence from India,” *Quarterly Journal of Economics*, 2010, *125* (4), 1727–1767.
- **and Frank Verboven**, “The Evolution of Price Dispersion in the European Car Market,” *Review of Economic Studies*, 2001, *68* (4), 811 – 848.
- Gong, Binlei and Robin C. Sickles**, “Resource Allocation in Multi-Divisional Multi-Product Firms,” *Journal of Productivity Analysis*, 2021, *55* (2), 47–70.
- Grieco, Paul L. E. and Ryan C. McDevitt**, “Productivity and Quality in Health Care: Evidence from the Dialysis Industry,” *Review of Economic Studies*, 2016, *84* (3), 1071–1105.
- 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 Steinar Strom, ed., *Econometrics and Economic Theory in the Twentieth Century: The Ragnar Frisch Centennial Symposium*, Cambridge University Press, 1998, pp. 169–203.
- Hall, Robert E.**, “The Relation between Price and Marginal Cost in U.S. Industry,” *Journal of Political Economy*, 1988, *96* (5), 921–47.
- Halpern, László, Miklós Koren, and Adam Szeidl**, “Imported Inputs and Productivity,” *American Economic Review*, 2015, *105* (12), 3660–3703.
- Harrigan, James, Ariell Reshef, and Farid Toubal**, “Techies, Trade, and Skill-Biased Productivity,” 2021. NBER working paper no. 25295, revised.
- Holtz-Eakin, Douglas, Whitney Newey, and Harvey S. Rosen**, “Estimating Vector Autoregressions with Panel Data,” *Econometrica*, Nov. 1988, *56* (6), 1371–1395.
- Hottman, Colin, Stephen J. Redding, and David E. Weinstein**, “Quantifying the Sources of Firm Heterogeneity,” *Quarterly Journal of Economics*, 2016, *131* (3), 1291–1364.
- Hsieh, Chang-Tai and Peter J. Klenow**, “Misallocation and Manufacturing TFP in China and India,” *Quarterly Journal of Economics*, 2009, *124* (4), pp. 1403–1448.
- Katayama, Hajime, Shihua Lu, and James R. Tybout**, “Firm-level Productivity Studies: Illusions and a Solution,” *International Journal of Industrial Organization*, May 2009, *27* (3), 403–413.
- Khandelwal, Amit K., Peter K. Schott, and Shang-Jin Wei**, “Trade Liberalization and Embedded Institutional Reform: Evidence from Chinese Exporters,” *American Economic Review*, 2013, *103* (6), 2169–95.
- Kleibergen, Frank**, “Testing Parameters in GMM Without Assuming That They Are Identified,” *Econometrica*, 2005, *73* (4), 1103–1123.
- **and Richard Paap**, “Generalized Reduced Rank Tests using the Singular Value Decomposition,” *Journal of Econometrics*, 2006, *133* (1), 97 – 126.
- Klette, Tor Jakob and Zvi Griliches**, “The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous,” *Journal of Applied Econometrics*, 1996, *11* (4), 343 – 361.
- Kraay, Aart**, “Weak Instruments in Growth Regressions: Implications for Recent Cross-Country Evidence on Inequality and Growth,” 2015. World Bank Policy Research Working Paper no. 7494.
- Kripfganz, Sebastian and Claudia Schwarz**, “Estimation of Linear Dynamic Panel Data Models with Time-Invariant Regressors,” *Journal of Applied Econometrics*, 2019, *34* (4), 526–546.
- Kugler, Maurice and Eric Verhoogen**, “Prices, Plant Size and Product Quality,” *Review of Economic Studies*, January 2012, *79* (1), 307–339.
- Lee, Yoonseok, Andrey Stoyanov, and Nikolay Zubanov**, “Olley and Pakes-style Production Function Estimators with Firm Fixed Effects,” *Oxford Bulletin of Economics and Statistics*, 2019, *81* (1), 79–97.
- Levinsohn, James and Amil Petrin**, “Estimating Production Functions Using Inputs to Control for Unobservables,” *Review of Economic Studies*, April 2003, *70*, 317–341.
- Maloney, William F and Jairo Nuñez Mendez**, “Measuring the Impact of Minimum Wages: Evidence from Latin America,” in James J. Heckman and Carmen Pagés, eds., *Law and Employment: Lessons from Latin America and the Caribbean*, University of Chicago Press, 2004.
- Marschak, Jacob and William H. Andrews**, “Random Simultaneous Equations and the Theory of Production,” *Econometrica*, 1944, *12* (3/4), 143–205.
- Melitz, Marc**, “Estimating Firm-level Productivity in Differentiated Product Industries,” 2000. Unpub. paper, Harvard University.

- Mondragón-Vélez, Camilo, Ximena Peña, and Daniel Wills**, “Labor Market Rigidities and Informality in Colombia,” *Economía*, 2010, *11* (1), 65–101.
- Montiel Olea, José Luis and Carolin Pflueger**, “A Robust Test for Weak Instruments,” *Journal of Business & Economic Statistics*, 2013, *31* (3), 358–369.
- Olley, G. Steven and Ariel Pakes**, “The Dynamics of Productivity in the Telecommunications Industry,” *Econometrica*, 1996, *64* (6), 1263–1297.
- Ornaghi, Carmine**, “Assessing the Effects of Measurement Errors on the Estimation of Production Functions,” *Journal of Applied Econometrics*, 2006, *21* (6), 879–891.
- Orr, Scott**, “Within-Firm Productivity Dispersion: Estimates and Implications,” *Journal of Political Economy*, 2022, *130* (11), 2771–2828.
- Park, Albert, Dean Yang, Xinzhen Shi, and Yuan Jiang**, “Exporting and Firm Performance: Chinese Exporters and the Asian Financial Crisis,” *Review of Economics and Statistics*, 2010, *92* (4), 822–842.
- Peter, Alessandra and Cian Ruane**, “The Aggregate Importance of Intermediate Input Substitutability,” 2020. Unpub. paper.
- Piveteau, Paul and Gabriel Smagghue**, “Estimating Firm Product Quality using Trade Data,” *Journal of International Economics*, 2019, *118*, 217–232.
- Redding, Stephen J. and David E. Weinstein**, “Measuring Aggregate Price Indices with Taste Shocks: Theory and Evidence for CES Preferences,” *Quarterly Journal of Economics*, 2020, *135* (1), 503–560.
- Roodman, David**, “How To Do xtabond2: An Introduction to Difference and System GMM in Stata,” *Stata Journal*, 2009, *9* (1), 86–136.
- Rovigatti, Gabriele and Vincenzo Mollisi**, “Theory and Practice of Total Factor Productivity Estimation: The Control Function Approach Using Stata,” *Stata Journal*, 2018, *18* (3), 618–662.
- Sanderson, Eleanor and Frank Windmeijer**, “A Weak Instrument F-test in Linear IV models with Multiple Endogenous Variables,” *Journal of Econometrics*, 2016, *190* (2), 212–221.
- Shenoy, Ajay**, “Estimating the Production Function when Firms are Constrained,” *Review of Economics and Statistics*, forthcoming.
- Smeets, Valerie and Frederic Warzynski**, “Estimating Productivity with Multi-product Firms, Pricing Heterogeneity and the Role of International Trade,” *Journal of International Economics*, 2013, *90* (2), 237 – 244.
- Söderbom, Måns and Francis Teal**, “Size and Efficiency in African Manufacturing Firms: Evidence from Firm-Level Panel Data,” *Journal of Development Economics*, 2004, *73* (1), 369–394.
- Stock, James H. and Jonathan H. Wright**, “GMM with Weak Identification,” *Econometrica*, 2000, *68* (5), 1055–1096.
- and **Motohiro Yogo**, “Testing for Weak Instruments in Linear IV Regression,” in “Identification and Inference for Econometric Models: A Festschrift in Honor of Thomas J. Rothenberg,” Cambridge UK: Cambridge University Press, 2005.
- Stock, James H, Jonathan H Wright, and Motohiro Yogo**, “A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments,” *Journal of Business and Economic Statistics*, 2002, *20* (4), 518–529.
- Tybout, James R.**, “Making Noisy Data Sing,” *Journal of Econometrics*, 1992, *53*, 25–44.
- Valmari, Nelli**, “Estimating Production Functions of Multiproduct Firms,” *Review of Economic Studies*, forthcoming.
- Verhoogen, Eric**, “Firm-Level Upgrading in Developing Countries,” *Journal of Economic Literature*, forthcoming.
- Windmeijer, Frank**, “A Finite Sample Correction for the Variance of Linear Efficient Two-Step GMM Estimators,” *Journal of Econometrics*, 2005, *126* (1), 25 – 51.
- Wooldridge, Jeffrey M.**, “On Estimating Firm-Level Production Functions Using Proxy Variables to Control for Unobservables,” *Economics Letters*, September 2009, *104* (3), 112–114.

Table 1. OLS and First Differences

	levels		first differences	
	(1)	(2)	(3)	(4)
A. Dependent variable: log sales				
log expenditures _{it}	0.675*** (0.027)	0.675*** (0.027)	0.488*** (0.050)	0.485*** (0.052)
log labor (ℓ_{it})	0.298*** (0.039)	0.296*** (0.040)	0.294*** (0.037)	0.288*** (0.036)
log capital (k_{it})	0.087*** (0.019)	0.087*** (0.019)	0.010 (0.019)	0.010 (0.018)
Year effects	N	Y	N	Y
R squared	0.926	0.927	0.335	0.339
B. Dependent variable: Sato-Vartia output index (\tilde{y}_{it}^{SV})				
Sato-Vartia materials index (\tilde{m}_{it}^{SV})	0.469*** (0.084)	0.468*** (0.085)	0.434*** (0.052)	0.428*** (0.053)
log labor (ℓ_{it})	0.357*** (0.109)	0.358*** (0.110)	0.283*** (0.046)	0.274*** (0.045)
log capital (k_{it})	0.196*** (0.045)	0.195*** (0.046)	0.013 (0.020)	0.015 (0.020)
Year effects	N	Y	N	Y
R squared	0.704	0.705	0.258	0.263

Notes: Baseline sample: N (observations) = 4,247, N (distinct firms) = 362 in all regressions. Columns 1-2 are in levels, columns 3-4 in first differences (between t-1 and t within the firm) for both independent and dependent variables. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 2. Differences (Step 1): First Stage

	$\Delta \tilde{m}_{it}^{SV}$	$\Delta \ell_{it}$	Δk_{it}	$\Delta \tilde{m}_{it}^{SV}$	$\Delta \ell_{it}$	Δk_{it}	$\Delta \tilde{m}_{it}^{SV}$	$\Delta \ell_{it}$	Δk_{it}
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
\tilde{m}_{it-2}^{SV}	-0.017*** (0.006)	0.013*** (0.004)	0.027*** (0.005)				-0.018*** (0.006)	0.012*** (0.004)	0.026*** (0.005)
ℓ_{it-2}	0.014 (0.009)	-0.030*** (0.007)	0.044*** (0.010)				0.013 (0.009)	-0.030*** (0.007)	0.044*** (0.010)
k_{it-2}	0.009 (0.006)	0.009** (0.004)	-0.048*** (0.007)	0.001 (0.003)	0.001 (0.002)	-0.010*** (0.003)	0.007 (0.006)	0.007* (0.004)	-0.050*** (0.007)
Δ pred. import price index ($\Delta \tilde{w}_{it}^{imp}$)				-0.248** (0.098)	-0.039 (0.064)	0.072 (0.103)	-0.255*** (0.098)	-0.045 (0.063)	0.118 (0.102)
Δ min. wage x “bite” (Δz_{it})				-1.366 (1.046)	-2.077*** (0.545)	-2.398*** (0.623)	-1.492 (1.049)	-2.062*** (0.549)	-1.991*** (0.628)
Year effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	4,247	4,247	4,247	4,247	4,247	4,247	4,247	4,247	4,247
R squared	0.024	0.035	0.038	0.024	0.032	0.014	0.026	0.039	0.041
F - statistic	3.424	7.703	21.018	3.814	6.882	6.08	3.522	7.437	13.921
F - SW	2.07	2.366	2.258	4.826	5.934	10.611	4.969	12.096	12.576
KP LM statistic (underidentification)		1.995			4.492			13.350	
KP LM p-value		0.158			0.034			0.004	
KP Wald F-statistic (weak insts.)		0.673			1.551			2.792	

Notes: Dependent variables at tops of columns. SW refers to Sanderson and Windmeijer (2016), KP to Kleibergen and Paap (2006). The F-statistic is the standard F for test that the coefficients on the excluded instruments (indicated at left) are zero. The KP statistics, LM test for under-identification and Wald F test for weak instruments, are for each IV estimator as a whole, and are not specific to Columns 2, 5, 8. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 3. Differences (Step 1): Second Stage

	Dep.var.: $\Delta \log$ output index ($\Delta \tilde{y}_{it}^{SV}$)			
	internal instruments only (1)	external instruments + k_{it-2} (2)	internal & external instruments (3)	internal & external instruments (LIML) (4)
$\Delta \log$ materials index ($\Delta \tilde{m}_{it}^{SV}$)	0.520 (0.487)	0.240 (0.528)	0.449** (0.192)	0.450** (0.195)
$\Delta \log$ labor ($\Delta \ell_{it}$)	0.485 (0.394)	0.688 (0.556)	0.471*** (0.176)	0.472*** (0.177)
$\Delta \log$ capital (Δk_{it})	-0.148 (0.196)	-0.166 (0.243)	-0.153 (0.133)	-0.154 (0.134)
Year effects	Y	Y	Y	Y
Observations	4,247	4,247	4,247	4,247
R-squared	0.224	0.185	0.237	0.237
Materials 90% Conf. Int.			[0.200 - 0.699]	[0.149 - 0.751]
Materials 95% Conf. Int.			[0.152 - 0.945]	[0.091 - 0.808]
Labor 90% Conf. Int.			[0.242 - 0.700]	[0.188 - 0.756]
Labor 95% Conf. Int.			[0.198 - 0.744]	[0.134 - 0.810]
Arellano-Bond AR(1) statistic	-2.890	-4.284	-4.304	-4.296
Arellano-Bond AR(1) p-value	0.004	0.000	0.000	0.000
Arellano-Bond AR(2) statistic	0.323	0.224	0.324	0.324
Arellano-Bond AR(2) p-value	0.746	0.823	0.746	0.746

Notes: Corresponding first-stage estimates are in Table 2: Column 1 here corresponds to Columns 1-3, Column 2 to Columns 4-6, Column 3 to Columns 7-9 of Table 2. Robust standard errors in parentheses. Confidence intervals are weak-instrument-robust, based on LC test of Andrews (2018), implemented by Stata twostepweakiv command. Arellano-Bond statistic and p-value test for serial correlation in residual, based on Arellano and Bond (1991). *10% level, **5% level, ***1% level.

Table 4. Levels (Step 2): First and Second Stages

A. First stage		
	Dep.var.: log capital (k_{it})	
	(1)	
Δ log capital, lagged (Δk_{it-1})	0.666*** (0.106)	
Year effects	Y	
N	4,247	
R squared	0.028	
KP LM statistic (underidentification)	43.269	
KP LM p-value	0.000	
KP Wald F-statistic (weak insts.)	39.453	
B. Second stage		
	Dep.var.: $\tilde{y}_{it}^{SV} - \hat{\beta}_m \tilde{m}_{it}^{SV} - \hat{\beta}_\ell \ell_{it}$	
	IV	OLS
	(1)	(2)
log capital k_{it}	0.114 [0.200]	0.151*** (0.020)
Year effects	Y	Y
N	4,247	4,247
R squared	0.077	0.082

Notes: Panel A reports the first stage and Panel B Column 1 the second stage of step 2 (levels step) of our two-step IV procedure. KP refers to Kleibergen and Paap (2006); see Section 2.5.1 and notes to Table 2 for details on KP tests. For comparison purposes, Panel B Column 2 reports OLS regression. Square brackets in Panel B Column 1 contain the corrected robust standard error; see Section 2.5.2 and supplementary materials available from the authors for explanation. Parentheses in Panel B Column 2 contain standard robust standard error. *10% level, **5% level, ***1% level.

Table 5. System GMM, Using Quantity Indexes

	log output index ($\Delta \tilde{y}_{it}^{SV}$)		
	(1)	(2)	(3)
\tilde{y}_{it-1}^{SV}	1.008*** (0.038)	0.991*** (0.036)	0.956*** (0.048)
\tilde{m}_{it}^{SV}	0.589*** (0.125)	0.659*** (0.097)	0.455*** (0.063)
\tilde{m}_{it-1}^{SV}	-0.596*** (0.124)	-0.662*** (0.105)	-0.401*** (0.069)
log labor _{it} (ℓ_{it})	0.310 (0.257)	0.203 (0.260)	0.292*** (0.110)
log labor _{it-1} (ℓ_{it-1})	-0.315 (0.238)	-0.216 (0.235)	-0.315*** (0.092)
log capital _{it-1} (k_{it})	0.070 (0.094)	0.139 (0.085)	0.106* (0.054)
log capital _{it-1} (k_{it-1})	-0.070 (0.089)	-0.112 (0.076)	-0.095** (0.049)
Observations	4,247	4,247	4,247
Lag limit	3	4	All
Hansen test	112.2	157.3	344.3
Hansen p-value	0.298	0.567	1.000

Notes: Table presents System GMM estimates (Blundell and Bond, 2000), using our quantity aggregates and the “two-step” procedure described in Roodman (2009), with initial weighting matrix defined in Doornik et al. (2012) and finite-sample correction from Windmeijer (2005). The Stata command is xtabond2 (Roodman, 2009), with options h(2), twostep, and robust. The difference equation includes lags from $t - 3$ in Column 1, from $t - 3$ and $t - 4$ in Column 2, and from $t - 3$ to firm’s initial year in Column 3. Weak-instrument diagnostics and the numbers of instruments are reported in supplementary materials available from the authors. The Hansen test of overidentifying restrictions is appropriate in the non-homoskedastic case, but should be interpreted with caution, as it is weakened by the presence of many instruments. See Section 5 for further details. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 6. Comparison to Other Methods

	OP	LP	GNR	GNR-MC	TSIV
	(1)	(2)	(3)	(4)	(5)
$\widehat{\beta}_m$ (materials)	0.669*** (0.028)	0.639*** (0.049)	0.406*** (0.009)	0.560*** (0.073)	0.450** (0.195)
$\widehat{\beta}_\ell$ (labor)	0.254*** (0.060)	0.291*** (0.045)	0.513*** (0.037)	0.597*** (0.085)	0.472*** (0.177)
$\widehat{\beta}_k$ (capital)	0.134*** (0.044)	0.108*** (0.041)	0.138*** (0.027)	0.272** (0.116)	0.114 (0.200)
N	1,933	4,247	4,247	4,247	4,247

Notes: Table presents estimates in our baseline sample from methods of Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), and Gandhi et al. (2020) (GNR), and an extension from GNR to allow for monopolistic competition (GNR-MC). OP and LP estimates generated by Stata command `prodest` (Rovigatti and Mollisi, 2018) including year effects, GNR estimates by our own coding of the GNR methods. See Section 5 for further details. Standard errors in parentheses from bootstraps with 50 replications. For comparison purposes, our estimates (TSIV) are reported in Column 6. *10% level, **5% level, ***1% level.

Table 7. Monte Carlo Simulation

	Dep. var.: log output quantity							
	OLS (1)	FD (2)	SysGMM (3)	OP (4)	LP (5)	GNR (6)	GNR-MC (7)	TSIV (8)
A. DGP1: Perfect Competition								
$\widehat{\beta}_m$ (materials)	0.720 (0.015)	0.783 (0.029)	0.730 (0.142)	0.692 (0.095)	0.720 (0.015)	0.650 (0.001)	0.650 (0.001)	0.646 (0.048)
$\widehat{\beta}_k$ (capital)	0.200 (0.011)	0.155 (0.021)	0.190 (0.104)	0.219 (0.075)	0.200 (0.011)	0.250 (0.001)	0.250 (0.001)	0.253 (0.037)
B. DGP2: Imperfect Competition								
$\widehat{\beta}_m$ (materials)	0.731 (0.026)	0.815 (0.039)	0.778 (0.216)	0.676 (0.040)	0.664 (0.068)	0.550 (0.001)	0.570 (0.031)	0.655 (0.066)
$\widehat{\beta}_k$ (capital)	0.212 (0.012)	0.173 (0.019)	0.190 (0.105)	0.238 (0.024)	0.243 (0.032)	0.298 (0.001)	0.221 (0.012)	0.248 (0.033)
C. DGP3: Imperfect Competition, Firm Effects								
$\widehat{\beta}_m$ (materials)	0.814 (0.020)	0.812 (0.038)	0.765 (0.197)	0.687 (0.042)	0.711 (0.084)	0.550 (0.001)	0.571 (0.033)	0.651 (0.070)
$\widehat{\beta}_k$ (capital)	0.173 (0.010)	0.174 (0.019)	0.194 (0.095)	0.232 (0.026)	0.221 (0.040)	0.298 (0.001)	0.221 (0.013)	0.249 (0.033)
D. DGP4: Imperfect Competition, Firm Effects, Input-Quality Differences								
$\widehat{\beta}_m$ (materials)	0.844 (0.019)	0.887 (0.034)	0.825 (0.198)	0.706 (0.039)	0.780 (0.213)	0.550 (0.001)	0.573 (0.034)	0.657 (0.059)
$\widehat{\beta}_k$ (capital)	0.159 (0.009)	0.139 (0.016)	0.169 (0.097)	0.224 (0.024)	0.188 (0.101)	0.298 (0.001)	0.222 (0.013)	0.246 (0.030)

Notes: Table presents Monte Carlo output-elasticity estimates for four data-generating processes (DGPs), all with serially uncorrelated productivity shocks. See Section 5.2 and Appendix D for details. The true elasticities are 0.65 for materials and 0.25 for capital. Physical quantities are used for output and materials, except for the first step of GNR and for GNR-MC, as explained in Appendix D. Table reports means and standard deviations of elasticity estimates for 100 samples of 15,000 observations each (except for OP, where we drop observations with zero investment). Means of standard errors estimated for each sample, averaged across samples, which to avoid clutter we do not report, are very similar to the reported standard deviations of the elasticity estimates.