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 October 2024

We thank Gianluca Antonecchia, David Atkin, Mons Chan, Jan De Loecker, 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 many seminar participants 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.

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 October 2024
JEL No. D24, L1, L65, O14

ABSTRACT

This paper uses output- and input-quantity information and external instruments for materials and labor choices to improve production-function estimates. In rich Colombian data on producers of rubber and plastic products, we construct the external instruments from exchange-rate movements and variation in the "bite" of the minimum wage. Under assumptions of constant elasticities of substitution among outputs and inputs within firms, we aggregate from the firm-product to the firm level and show how quality and variety choices may bias quantity-based estimators. We supplement the external instruments with internal instruments — lagged levels and differences of input choices — in a two-step approach, estimating a difference equation to recover the materials and labor coefficients and a levels equation to recover the capital coefficient. Our estimates imply markups that are 67-70% of those implied by standard proxy-variable methods. A simple Monte Carlo simulation illustrates the advantages of our approach in a controlled 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

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

Santiago Franco
Boston University
Department of Economics
270 Bay State Road, Office 405
Boston, MA 02215
safranco@bu.edu

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

"The future, it is hoped, 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." (Griliches and Mairesse, 1998, p. 198)

1 Introduction

A central task in the analysis of production functions is to estimate the elasticities of real output with respect to real inputs, unconfounded by unobserved differences in prices or productivity across firms. Such elasticities are key to calculating total factor productivity (TFP), the most common metric of firm performance, and to estimating markups in the influential method of Hall (1988) and De Loecker and Warzynski (2012). They thus play a pivotal role in a wide range of active economic debates, from misallocation (Hsieh and Klenow, 2009; Restuccia and Rogerson, 2013) to rising market power (De Loecker et al., 2020) and the declining labor share (Karabarbounis and Neiman, 2014; Autor et al., 2020).

Attempts to estimate such elasticities in differentiated-product industries face three important challenges, two well appreciated in the literature and one somewhat less so.[1] First, firms' input choices may respond to unobserved (to the econometrician) shocks to their productivity in a given period, generating "transmission bias" (Marschak and Andrews, 1944). Second, when one lacks information on physical quantities (as is usually the case), estimates using sector-level price indexes to deflate sales and input expenditures may be confounded by price variation at the firm level, generating what are often referred to as output- and input-price biases (Klette and Griliches, 1996; Foster et al., 2008, 2016; De Loecker and Goldberg, 2014). Bond et al. (2021) and Doraszelski and Jaumandreu (2023) have argued that De Loecker-Warzynski-type markup calculations using such estimates are problematic.

Less widely acknowledged is the third challenge: in differentiated-product industries, even if quantity information is available, regressions of physical output on physical inputs will not in general recover the desired elasticities. If product quality and variety are valued by consumers, they should be incorporated in our notion of real output, and if input quality and variety matter for real output, they should be incorporated in real inputs. But we will argue, in the spirit of Katayama et al. (2009) and others, that once one accepts these propositions, estimates using only physical units may be subject to *quality* and *variety biases*. To take a simple 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 understate the contribution of labor to real (i.e.

[1] For reviews, see Griliches and Mairesse (1998), Bartelsman and Doms (2000), Ackerberg et al. (2007), De Loecker and Goldberg (2014), De Loecker and Syverson (2021), and Section 2.2.1 of Verhoogen (2023).

quality-adjusted) output. Similar biases can result from changes in output or input variety or input quality. These quality and variety biases may be present even if one has a perfect proxy for unobserved productivity.

In this paper, we address all three challenges by bringing to bear two types of information that have seldom been exploited in the production-function literature and, to our knowledge, never in combination. The first is external instruments for materials and labor choices, capturing arguably exogenous determinants of input prices. Focusing on Colombian producers of rubber and plastic products, we construct these instruments using exchange-rate movements (in conjunction with customs data) and variation in the "bite" of minimum-wage changes, as described below. The second type of information is physical quantities of both outputs and inputs, which are reported on a consistent basis at a 7-digit product level in the Colombian manufacturing survey.

The idea that external instruments in general — and exogenous determinants of 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) quoted in the epigraph. 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.[2] Our contention is that this view is now too pessimistic: the expansion of the data frontier has made it possible to push forward the agenda that Griliches and Mairesse laid out a generation ago.

Although the Colombian data are as rich on the relevant dimensions as those from any other country, there remain two important data limitations. The first is that, as in almost all similar datasets, the product-level mapping between inputs and outputs is unobserved in multi-product firms. In contrast to recent work that has focused on single-product firms (e.g. De Loecker et al. (2016)) or has inferred allocations of inputs to outputs that would be made by optimizing firms (e.g. Orr (2022)), 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. Following common practice, we assume that (firm-level aggregate) materials, labor, and capital combine in a Cobb-Douglas function, which can be interpreted as a first-order approximation to any production function (Syverson, 2011).

An advantage of the within-firm CES structure is that it makes transparent how quality and

[2] For instance, in discussing Olley and Pakes (1996), Levinsohn and Petrin (2003) and related approaches, Ackerberg 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 related methodology (due to fewer auxiliary assumptions)." But they also note that "the premise of most of this literature is that such variables are either not available or not believed to be exogenous." See also Ackerberg et al. (2007, p. 4208) and Gandhi et al. (2020, Sec. VI.A).

variety differences may bias quantity-based output-elasticity estimates. Using existing results for CES aggregators, we show that changes in real firm-level aggregate output and materials can be expressed as sums of changes in observable quantity indexes and unobservable terms capturing quality and variety. These unobservable terms end up in the error term of a regression of output quantity on input quantities, and to the extent that they are correlated with input quantity choices they may generate the quality and variety biases mentioned above. Although the within-firm CES structure helps to elucidate these biases conceptually, we will show that our empirical results are robust to using other common aggregators.

The second important data limitation is that we do not have an external instrument capturing firm-level variation in the price or availability of capital. Given this constraint, we supplement the external instruments for materials and labor with internal instruments — lagged levels and differences of input choices — in the broad spirit of the System GMM approach of Arellano and Bover (1995) and Blundell and Bond (1998, 2000). From this literature we borrow the idea of combining an equation in differences, instrumented with lagged levels, with an equation in levels, instrumented with lagged differences. We depart from standard practice by estimating the equations in two steps, along the lines of a related exercise by Kripfganz and Schwarz (2019), and by including a more parsimonious set of internal instruments, which helps to address weak-instrument concerns. In the first step, we first-difference and include lagged input levels as instruments along with our external instruments. As has often been observed, first-differencing yields an implausibly low value of the capital coefficient, possibly because levels of capital stock are particularly poorly measured, even relative to (also noisy) investment,[3] and we treat the first-step 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 allows us to use lagged investment as an instrument for the level of capital. An advantage of the two-step approach is that the first-step materials and labor coefficients are robust to possible mis-specification of the levels equation; if one is just interested in estimating markups using the De Loecker and Warzynski (2012) method, one can stop at the first step. Our approach relies on a stronger restriction on the time-series evolution of productivity than is standard; this assumption is testable using standard methods, and we will not reject the null that the additional restriction holds.

Under assumptions that we argue are plausible in our setting, our two-step instrumental-variables (TSIV) estimator removes quality and variety biases as well as the familiar transmission bias. It yields reasonable point estimates: we find materials and labor coefficients of 0.45 and 0.47, respectively, and a capital coefficient of 0.11. Weak-instrument tests suggest reason for concern, so we report weak-instrument-robust confidence intervals using methods recently developed by Andrews (2016, 2018). The fact that constant returns to scale approximately hold, as one would generally expect (Bartelsman

[3] See e.g. Tybout (1992), Griliches and Mairesse (1998), Ornaghi (2006), Ackerberg et al. (2007), Ackerberg et al. (2015), and Collard-Wexler and De Loecker (2020).

and Doms, 2000), is reassuring.

We compare our estimates to those of other common methods — Olley and Pakes (1996, hereafter OP), Levinsohn and Petrin (2003, hereafter LP), Gandhi, Navarro and Rivers (2020, hereafter GNR), and Blundell and Bond (2000)[4] — in two ways. First, we implement these other methods in the Colombian data. Our confidence intervals are wide enough that the differences with standard estimators are generally not statistically significant, but the differences in point estimates carry potentially important economic implications. For example, if one were to treat materials as the flexible input in the De Loecker and Warzynski (2012) method, the markups calculated using our estimates would be just 67-70% of the levels calculated using OP or LP.[5] Second, we compare estimates across methods in 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 the other common estimators but that our estimator continues to perform well.

The next section briefly discusses related literature. Section 3 develops our econometric strategy. Section 4 describes the data we use and our motivation for focusing on producers of rubber and plastic products. Section 5 presents our estimates of output elasticities. Section 6 compares our coefficient estimates to those of other common estimation methods. Section 7 concludes.

2 Related Literature

In recent years, the dominant strategy in the production-function literature has been to construct a proxy for unobserved productivity by inverting either an investment-demand or a materials-demand equation (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Wooldridge, 2009; De Loecker, 2011; Doraszelski and Jaumandreu, 2013, 2018; Ackerberg et al., 2015; Eslava et al., 2023). This "proxy-variable" approach requires a monotonic relationship between the productivity term (assumed to be scalar) and investment or materials, conditional on other observables. The approach is unattractive in our setting for two reasons. First, as we argue below, our within-firm aggregation strategy requires firm-specific normalizations, which are incompatible with the scalar monotonicity assumption. Second, contracting frictions in credit and other input markets are commonly thought to be pervasive in developing countries and are likely to be present in our context. Any heterogeneity across firms in such frictions would also violate the required assumption (Shenoy, 2021). Heterogeneity in demand

[4]The method of Ackerberg, Caves and Frazer (2015, hereafter 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 be not be not directly comparable to ours and we do not include them in the comparison.

[5]Since in De Loecker and Warzynski (2012), the markup is calculated as simply the output elasticity to the share of revenues spent on the input, the ratio of markups can be calculated as the ratio of output elasticity estimates.

conditions also seems likely and would similarly invalidate the proxy-variable approach (Doraszelski and Jaumandreu, 2023).

As mentioned above, our approach is closer to the dynamic-panel-data literature, which does not require inverting an input-demand equation and more easily accommodates the firm effects (Chamberlain, 1982; Arellano and Bond, 1991; Arellano and Bover, 1995; Blundell and Bond, 1998, 2000). An advantage of using external instruments is that we are able to reduce our reliance on further and further lags of internal instruments, which may be only weakly related to current input choices.[6] While we do not avoid weak-instrument concerns completely, the external instruments help to strengthen our instrument set.

There is a small related literature on production-function estimation in multi-product firms using information at the firm-product level. As mentioned above, this literature has dealt in different ways with the lack of an observed mapping between inputs and outputs in multi-product firms. Papers that focus on single-product firms and (in some cases) implement a selection correction for the fact that they may not be representative include Foster et al. (2008), De Loecker et al. (2016), Forlani et al. (2023), and Blum et al. (2024). Papers that use estimates of demand elasticities and profit-maximization conditions to infer the allocation of inputs to outputs that would be implemented by optimizing firms include Orr (2022), Gong and Sickles (2021), Valmari (2023), and Caselli et al. (2023). Dhyne et al. (2022, 2023) develop an alternative strategy in which they relate output of a good to firm-level input usage and the output levels of other goods. Previous papers that aggregate from the sub-firm to the firm level, without explicitly considering quality and variety biases, include Eslava et al. (2004, 2013), Ornaghi (2006), Doraszelski and Jaumandreu (2013, 2018), Smeets and Warzynski (2013), Halpern et al. (2015), Garcia-Marin and Voigtländer (2019), 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 Danish data, 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. (2023), who

[6] For discussions in the setting of cross-country growth regressions, see Bazzi and Clemens (2013) and Kraay (2015). See also the overviews by Griliches and Mairesse (1998), Roodman (2009b) and Bun and Sarafidis (2015).

use quality-adjusted deflators constructed via joint estimation of production functions and demand.[7] Grieco and McDevitt (2016) and Atkin et al. (2019) take advantage of detailed product characteristics in particular sectors, outpatient dialysis centers in the US and the handwoven rug industry in Egypt, respectively.[8] Such direct measures of product quality are clearly very valuable for estimating firm performance, but unfortunately they are rarely available. We view our approach in this paper as being most useful in settings where product prices and quantities are available but detailed product characteristics are not. We are not aware of previous formalizations of variety biases in a production-function setting.

The number of previous papers attempting to use external instruments to improve production-function estimates is very small. An early paper by Nerlove (1963) exploited the exogeneity of factor prices in regulated public utilities. Syverson (2004), drawing on the input-output approach of Shea (1993), used employment in construction as an instrument for input usage in the ready-mix concrete industry. A recent paper by Ackerberg and De Loecker (2024) argues that, in imperfectly competitive settings, fixed inputs and productivity shocks of competing firms (if they are observed) can serve as instruments. We share the hope of Griliches and Mairesse (1998) that the future will bring further advances along these lines.

3 Econometric Strategy

This section first lays out the theoretical framework underlying CES aggregation on the demand side (Subsection 3.1) and production side (Subsection 3.2) and uses decompositions of the CES aggregates to rewrite the production function, which makes clear how quality and variety choices may bias standard estimates (Subsection 3.3). We then explain the construction of the external instruments (Subsection 3.4) and the assumptions required for our internal instruments to be valid (Subsection 3.5). Subsection 3.6 presents our two-step IV (TSIV) strategy. Derivations are in Appendix A.

3.1 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. Our approach is to impose constant elasticity of substitution (CES) of products within (but not across) firms. In particular, we assume that a representative

[7] De 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. (2023) (contemporaneously with this paper) also use CES aggregation, but in the context of joint estimation of production and demand functions that requires CES across as well as within firms.

[8] See also Hahn (2024) on the European car industry and Li et al. (2023) on Chinese steel.

consumer has the following utility function:

$$U_t = U(\widetilde{Y}_{1t}, \widetilde{Y}_{2t}, ..., \widetilde{Y}_{It}) \text{ where } \widetilde{Y}_{it} = \left[\sum_{j \in \Omega^y_{it}} (\varphi_{ijt} Y_{ijt})^{\frac{\sigma^y_i - 1}{\sigma^y_i}} \right]^{\frac{\sigma^y_i}{\sigma^y_i - 1}} \tag{1}$$

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, σ^y_i is the elasticity of substitution between products, specific to firm i, and Ω^y_{it} 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") per physical unit, which may reflect endogenous choices of the firm (e.g. physical attributes of goods) or external factors (e.g. exogenous fashion trends). Similar CES specifications have been used by Hottman et al. (2016) and others; we differ in that we do not require CES to hold across firms as well as within. We follow common practice and assume that $\sigma^y_i > 1$.[9]

The assumption of weak separability and the homotheticity of $\widetilde{Y}_{it}$ 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 $\widetilde{Y}_{it}$ and then choosing the $\widetilde{Y}_{it}$ (for $i = 1, 2, ..., I$) to maximize utility. Optimization in the first stage implies that the minimum price to purchase one unit of $\widetilde{Y}_{it}$ is:

$$\widetilde{P}_{it} = \left[\sum_{j \in \Omega^y_{it}} \left(\frac{P_{ijt}}{\varphi_{ijt}} \right)^{1-\sigma^y_i} \right]^{\frac{1}{1-\sigma^y_i}} \tag{2}$$

This is the price index that sets $\widetilde{P}_{it}\widetilde{Y}_{it} = R_{it}$, 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 product 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:

$$\widetilde{Y}_{it} = D_{it}(\widetilde{P}_{1t}, ..., \widetilde{P}_{It}, C_t) \text{ for } i = 1, 2, ..., I \tag{3}$$

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.

[9] Although the consumer optimization problem would remain well-behaved as long as $\sigma^y_i \in (0,1) \cup (1, \infty)$, the condition $\sigma^y_i > 1$ 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), $\sigma^y_i > 1$ 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.

The within-firm CES assumption allows us to decompose changes in the firm-specific price index in a particularly convenient way. Let $\Omega^{y*}_{it,t-1}$ be firm i's common outputs between $t-1$ and t (i.e. $\Omega^{y}_{it-1} \cap \Omega^{y}_{it}$), S^{y}_{ijt} be the consumer's period-t expenditure share on product j among all products produced by firm i, $S^{y*}_{ijt,t-1}$ be the period-t share among $(t, t-1)$ common goods, and $S^{y*}_{ijt-1,t}$ be the period-$(t-1)$ share among $(t, t-1)$ common goods.[10] 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:

$$\ln\left(\frac{\widetilde{P}_{it}}{\widetilde{P}_{it-1}}\right) = \sum_{j\in\Omega^{y*}_{it,t-1}} \psi^{y}_{ijt} \ln\left(\frac{P_{ijt}}{P_{ijt-1}}\right) - \sum_{j\in\Omega^{y*}_{it,t-1}} \psi^{y}_{ijt} \ln\left(\frac{\varphi_{ijt}}{\varphi_{ijt-1}}\right) - \frac{1}{\sigma^{y}_{i}-1} \ln\left(\frac{\chi^{y}_{it-1,t}}{\chi^{y}_{it,t-1}}\right) \tag{4}$$

where:

$$\psi^{y}_{ijt} = \frac{\left(\frac{S^{y*}_{ijt,t-1}-S^{y*}_{ijt-1,t}}{\ln S^{y*}_{ijt,t-1}-\ln S^{y*}_{ijt-1,t}}\right)}{\sum_{j\in\Omega^{y*}_{it,t-1}}\left(\frac{S^{y*}_{ijt,t-1}-S^{y*}_{ijt-1,t}}{\ln S^{y*}_{ijt,t-1}-\ln S^{y*}_{ijt-1,t}}\right)}, \quad \chi^{y}_{it,t-1} = \sum_{j\in\Omega^{y*}_{it,t-1}} S^{y}_{ijt}, \quad \chi^{y}_{it-1,t} = \sum_{j\in\Omega^{y*}_{it,t-1}} S^{y}_{ijt-1} \tag{5}$$

The first term on the right-hand side of (4) is (the log of) the familiar Sato-Vartia index (Sato, 1976; Vartia, 1976), a weighted average of product-specific price changes for common goods, with the "Sato-Vartia weights" ψ^{y}_{ijt}. This index is observable in our data. 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.[11] 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.[12] Although the σ^{y}_{i} term is unobservable, the $\chi^{y}_{it-1,t}$ and $\chi^{y}_{it,t-1}$ 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, $\widetilde{Y}_{it}$, can be expressed in a similar decomposition:

$$\ln\left(\frac{\widetilde{Y}_{it}}{\widetilde{Y}_{it-1}}\right) = \sum_{j\in\Omega^{y*}_{it,t-1}} \psi^{y}_{ijt} \ln\left(\frac{Y_{ijt}}{Y_{ijt-1}}\right) + \sum_{j\in\Omega^{y*}_{it,t-1}} \psi^{y}_{ijt} \ln\left(\frac{\varphi_{ijt}}{\varphi_{ijt-1}}\right) + \frac{\sigma^{y}_{i}}{\sigma^{y}_{i}-1} \ln\left(\frac{\chi^{y}_{it-1,t}}{\chi^{y}_{it,t-1}}\right) \tag{6}$$

---

[10] That is, $S^{y}_{ijt} = \frac{P_{ijt}Y_{ijt}}{R_{it}}$, and, for $j \in \Omega^{y*}_{it,t-1}$, $S^{y*}_{ijt,t-1} = \frac{P_{ijt}Y_{ijt}}{\sum_{j'\in\Omega^{y*}_{it,t-1}} P_{ij't}Y_{ij't}}$ and $S^{y*}_{ijt-1,t} = \frac{P_{ijt-1}Y_{ijt-1}}{\sum_{j'\in\Omega^{y*}_{it,t-1}} P_{ij't-1}Y_{ij't-1}}$.

[11] Redding 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.

[12] For example, if no goods are dropped from $t-1$ to t but new goods are introduced, then $\chi^{y}_{it-1,t} = 1 > \chi^{y}_{it,t-1}$, which, since $\sigma^{y}_{i} > 1$ 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.

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 $\sigma_i^y \to \infty$.[13] In that sense, our aggregation method is strictly more general than the common approach.

3.2 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:

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

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; it captures the contribution of one physical unit of the input to the input aggregate. $\widetilde{M}_{it}$. The assumption that the production function is Cobb-Douglas in capital, labor, and materials is standard in the literature. 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, 2023).[14]

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 $\widetilde{Y}_{it}$, which incorporates quality and variety as well as physical units. The ω_{it} term is an "ex ante" shock that firms observe 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

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

[14]As on the output side (see footnote 9), our approach remains applicable, although with somewhat less intuitive implications, as long as $\sigma_i^m \in (0,1) \cup (1,\infty)$.

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 3.5 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 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). Let W^m_{iht} be the price to firm i of purchasing input h in time t.[15] Optimization in the choice of input quantities implies that the cost of purchasing one unit of the materials aggregate, $\widetilde{M}_{it}$, is:

$$\widetilde{W}^m_{it} = \left[\sum_{h \in \Omega^m_{it}} \left(\frac{W^m_{iht}}{\alpha_{iht}} \right)^{1-\sigma^m_i} \right]^{\frac{1}{1-\sigma^m_i}} \tag{8}$$

This is the price index that sets $\widetilde{W}^m_{it} \widetilde{M}_{it} = E^m_{it}$, where E^m_{it} 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 $\Omega^{m*}_{it,t-1}$ be firm i's common inputs between $t-1$ and t (i.e. $\Omega^m_{it-1} \cap \Omega^m_{it}$), $E^{m*}_{it,t-1}$ be the firm's expenditures on $(t, t-1)$ common inputs, $S^m_{iht,t-1}$ be the firm's expenditure share on input h among all inputs purchased by firm i, $S^{m*}_{iht,t-1}$ be the period-t share among $(t, t-1)$ common inputs, and $S^{m*}_{iht,t-1}$ 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:

$$\ln\left(\frac{\widetilde{W}^m_{it}}{\widetilde{W}^m_{it-1}} \right) = \sum_{h \in \Omega^{m*}_{it,t-1}} \psi^m_{iht} \ln\left(\frac{W^m_{iht}}{W^m_{iht-1}} \right) - \sum_{h \in \Omega^{m*}_{it,t-1}} \psi^m_{iht} \ln\left(\frac{\alpha_{iht}}{\alpha_{iht-1}} \right) - \frac{1}{\sigma^m_i - 1} \ln\left(\frac{\chi^m_{it-1,t}}{\chi^m_{it,t-1}} \right) \tag{9}$$

[15] Although we allow variety and quality to be chosen endogenously, we assume that firms are price-takers for particular inputs, following De Loecker et al. (2016) and many others.

[16] That is, $S^m_{iht} = \frac{W^m_{iht} M_{iht}}{E^m_{it}}$, and, for $h \in \Omega^{m*}_{it,t-1}$, $S^{m*}_{iht,t-1} = \frac{W^m_{iht} M_{iht}}{\sum_{h' \in \Omega^{m*}_{it,t-1}} W^m_{ih't} M_{ih't}}$ and $S^{m*}_{iht-1,t} = \frac{W^m_{iht-1} M_{iht-1}}{\sum_{h' \in \Omega^{m*}_{it,t-1}} W^m_{ih't-1} M_{ih't-1}}$.

where:

$$\psi^m_{iht} = \frac{\left(\frac{S^{m*}_{iht,t-1} - S^{m*}_{iht-1,t}}{\ln S^{m*}_{iht,t-1} - \ln S^{m*}_{iht-1,t}}\right)}{\sum_{h \in \Omega^{m*}_{it,t-1}} \left(\frac{S^{m*}_{iht,t-1} - S^{m*}_{iht-1,t}}{\ln S^{m*}_{iht,t-1} - \ln S^{m*}_{iht-1,t}}\right)}, \quad \chi^m_{it,t-1} = \sum_{h \in \Omega^{m*}_{it,t-1}} S^m_{iht}, \quad \chi^m_{it-1,t} = \sum_{h \in \Omega^{m*}_{it,t-1}} S^m_{iht-1} \tag{10}$$

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]

$$\ln\left(\frac{\widetilde{M}_{it}}{\widetilde{M}_{it-1}}\right) = \sum_{h \in \Omega^{m*}_{it,t-1}} \psi^m_{iht} \ln\left(\frac{M_{iht}}{M_{iht-1}}\right) + \sum_{h \in \Omega^{m*}_{it,t-1}} \psi^m_{iht} \ln\left(\frac{\alpha_{iht}}{\alpha_{iht-1}}\right) + \frac{\sigma^m_i}{\sigma^m_i - 1} \ln\left(\frac{\chi^m_{it-1,t}}{\chi^m_{it,t-1}}\right) \tag{11}$$

3.3 Estimating Equations

To integrate the decompositions (6) and (11) into the production function, (7), it is convenient to restate them in levels. Let lower-case letters represent logs and $\triangle$ indicate changes 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:

$$\begin{aligned}
\widetilde{y}_{it} &= \underbrace{\widetilde{y}_{i0} + \sum_{\tau=1}^{t} \sum_{j \in \Omega^{y*}_{i\tau,\tau-1}} \psi^y_{ij\tau} \triangle y_{ij\tau}}_{=\widetilde{y}^{SV}_{it}} + \underbrace{\sum_{\tau=1}^{t} \sum_{j \in \Omega^{y*}_{i\tau,\tau-1}} \psi^y_{ij\tau} \ln\left(\frac{\varphi_{ij\tau}}{\varphi_{ij\tau-1}}\right)}_{=q^y_{it}} + \underbrace{\left(\frac{\sigma^y_i}{\sigma^y_i - 1}\right) \sum_{\tau=1}^{t} \ln\left(\frac{\chi^y_{i\tau-1,\tau}}{\chi^y_{i\tau,\tau-1}}\right)}_{=v^y_{it}} \qquad (12) \\
\widetilde{m}_{it} &= \underbrace{\widetilde{m}_{i0} + \sum_{\tau=1}^{t} \sum_{h \in \Omega^{m*}_{i\tau,\tau-1}} \psi^m_{ih\tau} \triangle m_{ih\tau}}_{=\widetilde{m}^{SV}_{it}} + \underbrace{\sum_{\tau=1}^{t} \sum_{h \in \Omega^{m*}_{i\tau,\tau-1}} \psi^m_{ih\tau} \ln\left(\frac{\alpha_{ih\tau}}{\alpha_{ih\tau-1}}\right)}_{=q^m_{it}} + \underbrace{\left(\frac{\sigma^m_i}{\sigma^m_i - 1}\right) \sum_{\tau=1}^{t} \ln\left(\frac{\chi^m_{i\tau-1,\tau}}{\chi^m_{i\tau,\tau-1}}\right)}_{=v^m_{it}}
\end{aligned}$$

where we define $\widetilde{y}^{SV}_{it}$, q^y_{it}, v^y_{it}, $\widetilde{m}^{SV}_{it}$, q^m_{it}, and v^m_{it}, as indicated by the underbraces. In defining variables in this way, we are setting the quality and variety terms q^y_{it}, v^y_{it}, q^m_{it}, and v^m_{it} to zero in the initial year and including the firm-specific normalizations as part of the Sato-Vartia quantity terms, $\widetilde{y}^{SV}_{it}$ and $\widetilde{m}^{SV}_{it}$. (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.)

[17] This approach again nests the standard approach of using expenditures deflated by a sector-level input price index as $\sigma^m_i \to \infty$; see footnote 13.

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

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

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 $\widetilde{m}_{it}^{SV}$, 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 $\widetilde{y}_{it}^{SV}$ and $\widetilde{m}_{it}^{SV}$ simplify to the physical quantities of the output and 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 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 input or output quality, if such shocks affect firms' input choices. For instance, if material input quality increases for exogenous reasons and this leads the firm to use more materials, there will be a positive bias in $\widehat{\beta}_m$. 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 $\widetilde{m}_{it}^{SV}$ and v_{it}^m and between $\widetilde{m}_{it}^{SV}$ and v_{it}^y, generating offsetting biases with ambiguous net effects. As noted above, 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 (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; Ackerberg et al., 2015) Here the firm-specific initial-year quantity terms, $\widetilde{y}_{i0}$ and $\widetilde{m}_{i0}$, in (12), which are embedded in $\widetilde{y}_{it}^{SV}$ and $\widetilde{m}_{it}^{SV}$ in (13), introduce additional heterogeneity across firms that would violate the scalar monotonicity assumption.[18] We included a firm effect at the outset in the production function, (7), but a firm effect would also be needed to absorb these firm-specific normalizations. Recent work has shown that proxy-variable strategies using choices of static inputs to construct the proxy can

[18]These quantity terms can also be thought of as reflecting normalizations of firm-specific output and input price indexes.

accommodate firm effects in productivity (see Lee et al. (2019) and Ackerberg (2021)), but time-invariant firm-level differences in prices or quantities would still be problematic in this approach.[19] We pursue an approach more in the spirit of the panel-data literature, in part because it can more easily accommodate firm fixed effects.

In differences, (13) becomes:

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

We refer to (14) as our difference equation.[20] Note that, given the definitions of v_{it}^m, v_{it}^y, q_{it}^m, and q_{it}^y in (12), $\triangle u_{it}$ can be rewritten as:

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

3.4 External Instruments

To estimate (14), we first seek external instruments that are correlated with $\triangle\widetilde{m}_{it}^{SV}$ and $\triangle\ell_{it}$ and uncorrelated with the error term, $\triangle u_{it}$. To construct the materials-price instrument, we take advantage of trade-transactions data merged with the Colombian manufacturing survey. We 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.[21] We then use information on the product composition of each firm's imports to aggregate the predictions to the firm-year level. The datasets are described in Section 4 below.

We begin by defining real exchange rates (RERs) as:

$$RER_{ot} = NER_{ot}\left(\frac{CPI_{ot}}{CPI_{Col,t}}\right) \tag{16}$$

where o indexes import origins, NER_{ot} is the nominal exchange rate (Colombian pesos/foreign cur-

[19] This issue is separate from the (valid) concerns that firms may face heterogeneous contracting constraints in input markets or heterogeneous demand conditions, which might also break the monotonic relationship between productivity and materials or investment demand (Shenoy, 2021; Doraszelski and Jaumandreu, 2023).

[20] A "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.

[21] Exchange-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.

rency), 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 could arguably be considered capital rather than material imports; we also exclude petroleum and other mineral fuels.[22] 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:

$$\triangle \overline{rer}_{nt,-i} = \sum_{o \in \mathcal{O}} \zeta_{ont-1,-i} \triangle \ln(RER_{ot}), \quad \text{where } \zeta_{ont-1,-i} = \frac{\mathcal{I}_{ont-1,-i}}{\sum_{o \in \mathcal{O}} \mathcal{I}_{ont-1,-i}} \tag{17}$$

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. We then use these product-level average real-exchange-rate changes to predict import price changes at the product-year level, using the regression:

$$\triangle w^{imp}_{nt,-i} = \gamma_{st,-i} \triangle \overline{rer}_{nt,-i} + \rho_{st,-i} + \check{v}_{nt,-i} \tag{18}$$

where $\triangle w^{imp}_{nt,-i}$ 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,[23] $\rho_{st,-i}$ is a sector-year effect, and $\check{v}_{nt,-i}$ is a product-year-level disturbance. In our preferred specification, s indexes two digit trade sectors.[24] 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) and recover the predicted values, $\triangle \widehat{w}^{imp}_{nt,-i}$. The advantage of using the predicted values, $\triangle \widehat{w}^{imp}_{nt,-i}$, as opposed to $\triangle w^{imp}_{nt,-i}$, 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 $\check{v}_{nt,-i}$), 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 $\triangle w^{imp}_{nt,-i}$.

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:

$$\triangle \widehat{\overline{w}}^{imp}_{it} = \sum_{n \in \mathcal{N}} \theta_{int-2} \triangle \widehat{w}^{imp}_{nt,-i}, \quad \text{where } \theta_{int-2} = \frac{\mathcal{I}_{int-2}}{\sum_{n \in \mathcal{N}} \mathcal{I}_{int-2}} \tag{19}$$

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

[23] That is, $\triangle w^{imp}_{nt,-i} = \sum_{o \in \mathcal{O}} \zeta_{ont-1,-i} \triangle w^{imp}_{ont,-i}$, where $\zeta_{ont-1,-i}$ is defined as in (17).

[24] In principle, we could include lags of the average real-exchange-rate changes in (18). 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.

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 $\triangle\widehat{\widetilde{w}}_{it}^{imp} = 0$.[25] This average predicted import price change at the firm level, $\triangle\widehat{\widetilde{w}}_{it}^{imp}$, is our external instrument for $\triangle\widetilde{m}_{it}^{SV}$ in (14). To be clear, the variation across firms in this instrument is primarily due to differences in the mix of imported inputs used by each firm, not to the fact that we leave out one firm at a time when predicting import prices.[26]

In order for $\triangle\widehat{\widetilde{w}}_{it}^{imp}$ to be a valid instrument, 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, $\triangle u_{it}$ in (14)-(15). That the exchange-rate movements in $\widehat{w}_{nt,-i}^{imp}$ are uncorrelated with the ex-ante and ex-post productivity shocks, $\triangle\omega_{it}$ and $\triangle\epsilon_{it}$, seems uncontroversial. Under the assumptions we spell out in the next section, import composition in $t-2$ is also uncorrelated with these shocks. One might be more worried that the instrument is correlated with the input and output quality and variety terms in $\triangle u_{it}$, for instance if the quality of imports varies systematically by origin and RER movements lead firms to source from different origins. To mitigate this concern, we have chosen to focus on sectors — plastic and rubber products — in which the 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 discuss the selection of sectors in more detail in Section 4.4 below.

To construct an external instrument for labor, we exploit the facts that the minimum wage in Colombia is high relative to the wage distribution (above 90% of the median) and rose sharply over our sample period, especially in 1994-1999 and 2003-2009. (See Section 4.3 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:

$$B_{it} = \frac{MW_t}{W_{it}^{\ell}} \tag{20}$$

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, $B_{it} \leq 1$; the closer the firm average wage is to the national minimum wage, the larger is B_{it}. We

[25] If 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 impracticable.

[26] Exchange-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 Appendix C.4.

interact this measure of bite with the change in the national minimum wage, using bite from $t-2$:

$$\triangle z_{it} = B_{it-2} * \triangle \ln(MW_t) \tag{21}$$

We maintain the assumption that $\triangle z_{it}$ is uncorrelated with the differenced error term, $\triangle u_{it}$, conditional on the (differenced) year effect, $\triangle \xi_t$, 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).

3.5 Internal Instruments

We supplement the external instruments with a parsimonious set of internal instruments. This subsection lays out a set of timing assumptions under which lagged input levels from $t-2$ and earlier are valid instruments in our estimating equation in differences, (14). 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.

Consider the ex-ante and ex-post productivity shocks, ω_{it} and ϵ_{it}. Let $\mathscr{I}_{it}$ 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 effect, η_i, the year effect, ξ_t, and the realization of ω_{it} (but not of ϵ_{it}) for period t. We assume:

$$\mathbb{E}(\epsilon_{it}|\mathscr{I}_{it}) = 0 \tag{22}$$

$$\mathbb{E}(\omega_{it}|\mathscr{I}_{it-1}) = 0 \tag{23}$$

Since $\epsilon_{it-1} \in \mathscr{I}_{it}$ and $\omega_{it-1} \in \mathscr{I}_{it-1}$, (22) and (23) imply that both ϵ_{it} and ω_{it} are serially independent, conditional on the firm fixed effect. The former assumption is standard. While the latter assumption is restrictive — the literature usually assumes assumes a first-order Markov process in the ex-ante productivity term (Olley and Pakes, 1996; Levinsohn and Petrin, 2003; Ackerberg 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 the firm fixed effect and contemporaneous variables — in particular the ex-ante productivity shock, ω_{it}, and possibly exogenous shifters that are serially independent conditional on the firm fixed effects. Formally, let $\vec{\Omega}_{it}^m$ and $\vec{\Omega}_{it}^y$ 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 $\vec{\alpha}_{it}$ and $\vec{\varphi}_{it}$ be vectors of qualities for inputs in Ω^m_{it} and outputs in Ω^y_{it}. Let $\vec{\Gamma}^{vm}_{it}$, $\vec{\Gamma}^{qm}_{it}$, $\vec{\Gamma}^{vy}_{it}$, and $\vec{\Gamma}^{qy}_{it}$ be vectors of exogenous, potentially firm-specific shifters that affect input and output variety and quality. We assume that these shifters are serially independent conditional on the firm effects, η_i: $\mathbb{E}(\vec{\Gamma}^{vm}_{it}|\vec{\Gamma}^{vm}_{it-1},\eta_i) = \mathbb{E}(\vec{\Gamma}^{vy}_{it}|\vec{\Gamma}^{vy}_{it-1},\eta_i) = \mathbb{E}(\vec{\Gamma}^{qm}_{it}|\vec{\Gamma}^{qm}_{it-1},\eta_i) = \mathbb{E}(\vec{\Gamma}^{qy}_{it}|\vec{\Gamma}^{qy}_{it-1},\eta_i) = \vec{0}$. Our assumption on variety and quality choices can then be written as: $\vec{\Omega}^m_{it} = F^m(\omega_{it},\eta_i,\vec{\Gamma}^{vm}_{it})$, $\vec{\Omega}^y_{it} = F^y(\omega_{it},\eta_i\vec{\Gamma}^{vy}_{it})$, $\vec{\alpha}_{it} = G^m(\omega_{it},\eta_i,\vec{\Gamma}^{qm}_{it})$, and $\vec{\varphi}_{it} = G^y(\omega_{it},\eta_i,\vec{\Gamma}^{qy}_{it})$. Given that by assumption ω_{it} and the exogenous shifters are serially independent conditional on η_i, firms' variety and quality choices are also conditionally serially independent.

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 in logs as the sum of a firm-level log average input price, $\overline{w}^m_{it}$, and a firm-input-specific term, ι_{iht}, which is (conditionally) serially independent:

$$w^m_{iht} = \overline{w}^m_{it} + \iota_{iht} \quad \text{where } \mathbb{E}(\iota_{iht}|\mathscr{I}_{it-1}) = 0 \tag{24}$$

The key feature of (24) is that any serial correlation in input prices is limited to the log firm-average price term, $\overline{w}^m_{it}$. (Note that, given our other assumptions, we will need $\overline{w}^m_{it}$ 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 analysis is complicated by the fact that firms endogenously choose output prices, which affect revenue shares, which enter the quality and variety terms, q^y_{it} and v^y_{it}. To solve explicitly for such output-price choices requires a specification of the micro-foundation for the firm-level production function, (7). A sufficient condition for our approach to be valid is that, like input prices, log output prices can be written as the sum of a potentially serially correlated firm-level term and a (conditionally) serially independent firm-product-level term:

$$p_{ijt} = \Lambda_{it} + \varsigma_{ijt} \quad \text{where } \mathbb{E}(\varsigma_{ijt}|\mathscr{I}_{it-1}) = 0 \tag{25}$$

Appendix A.3 provides a micro-foundation for our production function that is consistent with (25). 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 this is not the only possible micro-foundation for our production function (7) and that our approach will be valid as long as (25) 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 (24) and (25) drop out of the expressions for the input-expenditure shares, S^m_{iht}, and output-revenue shares, S^y_{ijt}, with the consequence that the shares are serially independent conditional on the firm fixed effects. Given the definitions of the Sato-Vartia weights, ψ^m_{iht} and ψ^y_{iht}, and the variety terms, $\chi^m_{it,t-1}$, $\chi^m_{it-1,t}$, $\chi^y_{it,t-1}$, and

$\chi^y_{it-1,t}$ in (5) and (10), it follows from (15) that $\triangle v^m_{it}$, $\triangle v^y_{it}$, $\triangle q^m_{it}$, and $\triangle q^m_{it}$ are $MA(1)$, i.e. display at most one period of serial correlation. Together with (22)-(23), this implies that input choices from $t-2$ and earlier are uncorrelated with $\triangle u_{it}$.

We recognize that the restrictions laid out in this section — in particular, the conditional serial independence of the ex-ante productivity shocks, ω_{it} — are stronger than those typically imposed in the literature. At the same time, the key implication of this set of assumptions — that $\triangle u_{it}$ is $MA(1)$ — is testable using the standard approach of Arellano and Bond (1991). We will see below that we do not reject the hypothesis of no correlation between $\triangle u_{it}$ and $\triangle u_{it-2}$, which increases our confidence in the assumptions. It is also worth emphasizing that while imposing conditional serial independence of the external factors that affect quality and variety decisions is also restrictive, it is less restrictive than the common practice of assuming away quality and variety differences entirely.

3.6 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. To be clear on terminology, we use the word *step* to refer to the estimation in differences or levels, and reserve the word *stage* for the two stages of the IV estimation in each step.

In the first step, we estimate the difference equation, (14), using the external instruments described in Section 3.4 and lagged levels from $t-2$ as internal instruments. 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; Ackerberg 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 24), 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 the data to estimate β_k well in first-differences. At the same time, we would not be justified in simply dropping $\triangle k_{it}$ 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 $\triangle k_{it}$ in the first step and treat β_k as a nuisance parameter.

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.

In the second step, we insert the first-step estimates of β_m and β_ℓ into the levels equation and use the lagged difference of capital as an instrument for the level. Our levels equation, (13), can be rewritten as:

$$\widetilde{y}_{it}^{SV} - \widehat{\beta}_m \widetilde{m}_{it}^{SV} - \widehat{\beta}_\ell \ell_{it} = \beta_k k_{it} + \xi_t + \breve{u}_{it} \tag{26}$$

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

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

For lagged differences to be valid instruments in this equation, we need an additional assumption. Following the System GMM literature, we assume that differences have a zero expectation conditional on the firm effect, η_i:

$$\mathbb{E}(\triangle k_{it} | \eta_i) = 0 \tag{27}$$

This assumption is often referred to as a "mean stationarity" assumption or a "constant correlated effects" assumption since it implies that $\mathbb{E}(\triangle k_{it}\eta_i) = c_i$ for some constant c_i (Bun and Sarafidis, 2015). Although the assumption rules out correlation between a firm's time-invariant productivity and the evolution of its capital stock over time, it allows investment to be a function of current and past shocks to productivity (as well as input prices). We recognize that this assumption is restrictive; it is also the standard assumption in the System GMM approach.[27] Appendix A.5 shows that, under this assumption and those in Section 3.5, the lagged difference in capital, $\triangle k_{it-1}$, is uncorrelated with $\breve{u}_{it}$ and hence is a valid instrument.

To estimate the levels equation, we also need to take a stand on the firm-specific normalizations, $\widetilde{m}_{i0}$ and $\widetilde{y}_{i0}$, in (12). This amounts to choosing a base year for the firm-specific output and input price indexes, $\widetilde{P}_{it}$ and $\widetilde{W}_{it}^m$. Here we assume that these indexes are equal to unity in the first year

[27] Refer to equation 3.5 in Blundell and Bond (2000) or equation 2.6 in Bun and Sarafidis (2015).

that a firm appears in our data. In logs, since $r_{it} = \widetilde{y}_{it} + \widetilde{p}_{it}$ in every period, setting $\widetilde{p}_{i0} = 0$ implies $\widetilde{y}_{i0} = r_{i0}$ (where 0 refers to the initial year for the firm). Similarly, on the input side, if we let e_{it} be log expenditures, then $e_{it} = \widetilde{m}_{it} + \widetilde{w}^m_{it}$ and setting $\widetilde{w}^m_{i0} = 0$ implies $\widetilde{m}_{i0} = e_{i0}$.

Although the first-step estimation errors in $\widehat{\beta}_m$ and $\widehat{\beta}_\ell$ show up in $\breve{u}_{it}$, 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 $\triangle k_{it-1}$ and $\widetilde{m}^{SV}_{it}$ or ℓ_{it}, a correction needs to be applied to the standard error for $\widehat{\beta}_k$ (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 estimation of the two equations. 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, (27).

4 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.

4.1 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. In the sectors we focus on, nearly all firms have just one plant. We focus on data from the period 1994-2009.[28] 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.[29] 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}

[28]Prior 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. The end of the study period, 2009, is determined by the availability of firm-level links between the EAM survey and customs records.

[29]The 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.

is the value of product j 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 Ackerberg et al. (2015), that the exclusion restrictions for input-price instruments are likely to be violated if one observes only a subset of inputs.

4.2 Customs Records and Exchange Rates

The customs data contain information from declarations 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 trade categories based on Harmonized System), year, and country of origin (for imports) or destination (for exports). The data have been merged with the EAM manufacturing data using firm identifiers. 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. Further details on the customs data can be found in Appendix B.2 and on exchange rates in Appendix B.3.

4.3 Minimum Wage

Despite wide variation in local labor market conditions, Colombia has a single national minimum wage. It increased significantly in real terms over our study period. 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. The real value of the minimum wage continued to increase after the ruling, rising by 23% during our study period, as illustrated in Appendix Figure A1. It remained above 90% of the median wage throughout the period — one of the highest such ratios in Latin America (Mondragón-Vélez et al., 2010). To illustrate the "bite" of the minimum wage, Appendix Figure A2 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*. (Details in Appendix B.4.) There was extensive bunching of wages at the minimum in 1998 (solid vertical line), and a large share of manufacturing workers was directly affected by the increase in

1999 (dashed vertical line). 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).

4.4 Choice of Subsectors and Descriptive Statistics

Our approach is most applicable in industries that meet several criteria. First, it is most useful 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 approach 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 to increase precision. 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 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 7-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 in 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 7-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 delivery 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 firms to 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. Although firm-product-level input and output quantities are available over the entire study period, 1996-2009, the EAM itself contains information on imports and exports only in 2000-2009, so we focus on this period in Panel B. The overall message is that the two subsectors are comparable on many dimensions. It is noteworthy that the share of single-product firms is small — just 15% the pooled sample. This suggests a concern with methods that seek to infer the mapping between inputs and outputs in multi-product firms from the observed mapping in single-product firms: the single-product firms seem likely to be non-representative in this setting.

5 Results

This section reports the results of the estimation strategy laid out in Section 3. For comparison purposes, we begin by presenting "naive" OLS and first-difference (FD) results, and then move on to our two-step IV (TSIV) approach.

5.1 "Naive" OLS and FD Estimators

Panel A of Table 1 presents estimates using sales as the measure of output and material expenditures as the measure of input use (deflated using sector-level deflators). The OLS estimates appear in Columns 1 and 2, without and with year effects respectively.[30] Columns 3 and 4 report first-difference (FD) estimates. 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 approach.

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 $\widetilde{y}_{i0}$ and $\widetilde{m}_{i0}$ discussed in Section 3.6 above, using each firm's first year in the unbalanced panel as the base year for the firm-specific output and input deflators. Comparing Panels A and B, we see significant differences in the OLS estimates — in particular, using the quantity indexes reduces the materials coefficient and raises the capital coefficient — but the FD estimates are quite similar across panels.

5.2 Differences (Step 1) Results

In the first 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 the internal instruments — lagged levels of inputs from period $t-2$. The coefficient estimates are plausible, with lagged levels negatively associated with current changes. But how strong are the instruments? Testing for weak instruments is complicated in this setting by the presence of multiple endogenous covariates and the potential for heteroskedastic errors. This is a frontier area of econometric theory and there is no consensus on the right diagnostic tests for such cases.[31] Two tests are commonly reported in practice. Sanderson and Windmeijer (2016) propose an improved version of a test from Angrist and Pischke (2009), which is appropriate for inference on each of multiple endogenous regressors.[32] Also commonly reported is the Kleibergen and Paap (2006) Wald statistic, an analogue of the Cragg and Donald (1993) statistic applicable in non-homoskedastic

[30] The shares of revenues paid to each input are often used as simple estimators of output elasticities (Syverson, 2011). In our case, the revenue shares of materials, labor and capital are .67, .21, and .07, very close to the OLS estimates in levels in Panel A. We note that the assumptions under which these revenue shares reflect true output elasticities — in particular, assumptions of perfect competition and constant returns to scale — are quite restrictive.

[31] In 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.

[32] The 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.

settings. Using these diagnostics, there is evidence that the internal 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 under-identification, the KP Wald statistic for weak instruments is below 1.[33] This weak-instrument issue is not resolved by including further lags as instruments in a GMM estimator. (See Appendix C.1.)

Columns 4-6 of Table 2 report the first-stage estimates including the two external instruments — the predicted change in import price, $\triangle\widehat{\overline{w}}_{int}^{imp}$ from (19), and the minimum wage instrument, $\triangle z_{it}$ from (21) — and one internal instrument, the lagged level of capital from $t-2$.[34] 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 instruments in Table 2. In the first two columns, the coefficients on materials and labor are imprecisely estimated and differ markedly, 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 $\triangle\ell_{it}$ and $\triangle q_{it}^{y}$ in (14), generating a negative bias in OLS and FD estimates of β_ℓ, which our approach would correct.

[33]Although 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 the statistic is at a level that would typically raise concerns among practitioners.

[34]Appendix C.2 reports the auxiliary leave-one-out regressions of import prices on RER movements we run in constructing the import-price instrument.

As previewed above, the capital coefficient is implausibly low in this specification. The point 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 the weak-instrument concern, 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.

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 $\triangle u_{it}$ and $\triangle u_{it-1}$, unsurprisingly, but we fail to reject the null of no correlation between $\triangle u_{it}$ and $\triangle u_{it-2}$. As discussed in Section 3.5 above, this is consistent with the assumptions required for our internal instruments to be valid in this context.

5.3 Levels (Step 2) Results

We now turn to the second step of our TSIV procedure, in levels. We estimate equation (26), 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 $\triangle k_{it-1}$ as the instrument for k_{it}. Weakness of the instrument is not a concern here: the Kleibergen-Paap Wald statistic is above 39.[35] 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 3.6 above. The corrected standard error does not allow us to

[35]The Kleibergen-Paap Wald statistic and the Sanderson-Windmeijer F statistic coincide in cases with a single endogenous covariate.

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 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).

5.4 Robustness

Appendix Sections C.3-C.5 report 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.

6 Comparison to Other Methods

In this section, we compare our output-elasticity estimates to those of other commonly used methods in the Colombian data (Subsection 6.1) and then explore the performance of the various methods in a simple Monte Carlo simulation (Subsection 6.2).

6.1 In Colombian Data

To compare our estimates to those of System GMM (Blundell and Bond, 2000), 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 (2009a).[36] The coefficients on the contemporaneous log materials quantity index, log labor, and log capital are estimates of the Cobb-Douglas output elasticities, corresponding to our β_m, β_ℓ, and β_k.[37] 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, 2009a). The materials (0.455) and capital (0.106) coefficients from the "all lags" specification (Column 3) are quite similar to our TSIV estimates (0.449

[36]In particular, we use the Stata xtabond2 command of Roodman (2009a) 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.

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

and 0.114, respectively). The main difference is in the labor coefficients, where the System GMM coefficient (0.292) is smaller than ours (0.471). Additional details are in Appendix C.6.

For the other methods we consider, we rely on revenues and expenditures deflated at the sector level, the variables that are available in standard datasets.[38] Table 6 reports estimates from the methods of Olley and Pakes (1996) (OP), Levinsohn and Petrin (2003) (LP), Gandhi, Navarro and Rivers (2020) (GNR), and an extension of GNR that allows for monopolistic competition in output markets, which we label GNR-MC.[39] (As mentioned above, Ackerberg, Caves and Frazer (2015) (ACF) recommend that their method only be used with value-added production functions, not gross output functions, hence coefficient estimates from their method would not be not 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 higher and for the point estimates for labor are 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.

Although the confidence intervals are sufficiently wide that the differences in coefficient are generally not statistically significant, the differences in point estimates carry potentially important economic implications. As one illustration, consider the markups that would be constructed using the De Loecker and Warzynski (2012) method in our data, calculated as the ratio of the output elasticity for that input to the share of revenues spent on that input. If materials are assumed to be the flexible input, as in De Loecker and Warzynski (2012), then ratios of markup estimates from the various methods can be calculated as the ratio of materials coefficient estimates. Using the estimates from Table 6, markups based on our TSIV estimates would be 67% of those based on OP (0.45/0.67), 70% of those based on LP, 111% of those based on GNR, 90% of those based on GNR-MC, and nearly identical to those based on Blundell and Bond (2000) using all available lags (in Column 3 of Table 5).) If these patterns were to hold at more aggregate level, they would matter greatly for ongoing macroeconomic debates about the role of concentration and market power.

[38] It is natural to ask whether one could use the quantity indexes with these other methods, but our aggregation strategy requires firm-specific normalizations that in general would be incompatible with the scalar monotonicity assumption required by proxy-variable approaches.

[39] For OP and LP, 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.

6.2 Monte Carlo Simulation

While the comparison in the previous subsection reveals several notable 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 estimator 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.

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 GNR. 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 independent 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 5,000 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).

Table 7 presents the results. 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.

In Column 1, we see that OLS performs poorly across DGPs, rejecting the true values of β_m and β_k at a high level of confidence. Under DGP1, the issue is transmission bias — the transmission of productivity shocks into contemporaneous materials choices (which are correlated with capital stock – hence the capital coefficient is also biased). Interestingly, moving from DGP1 to DGP2 reduces the transmission bias somewhat, because imperfect competition dampens the response of firm output — and hence of input quantity – to productivity shocks. Moving from DGP2 to DGP3, the addition of firm effects in productivity increases the variance of productivity and the transmission bias is correspondingly larger. In DGP4, the addition of input-quality differences leads to a positive input-quality bias in $\widehat{\beta}_m$.

In Column 2, the first-differences (FD) estimator handles the firm effects in productivity better, unsurprisingly, but generally performs even worse than OLS, because the differencing substantially exacerbates the transmission bias.

In Column 3, the SysGMM estimator yields larger standard errors than OLS or FD, due in part to the weakness of the many included internal instruments, and although the point estimates for DGP1-DGP2 are similar to OLS, the true values are within a 95% confidence interval. SysGMM handles the inclusion of firm effects in productivity better than OLS, and little bias is introduced moving from DGP2 to DGP3. But in DGP4, there is a positive input-quality bias as in OLS and FD; the true value of β_m is rejected at the 95% level in Panel D.

The OP estimator, reported in Column 4, is arguably not well suited to the context of the simulation, because it relies heavily on serial correlation in ex-ante productivity. In DGP1, with no firm effects and serially independent productivity shocks, high productivity today does not indicate high productivity tomorrow, there is no reason to invest more in response to a high productivity shock today, and hence investment does not proxy for the unobserved ex-ante productivity term. As a result, including the proxy has little effect on the estimates and they are similar to OLS.

Considering the LP estimator in Column 5, we note that static materials choices (unlike investment) do proxy for productivity in this context. But initial differences in input prices and hence output prices represent an additional dimension of heterogeneity across firms, violating the scalar monotonicity assumption. Another under-appreciated issue is the sensitivity of estimates in proxy-variable methods to the choice of starting values, which has been noted for instance by Rovigatti and Mollisi (2018) and Norris-Keiller et al. (2024).[40] Following common practice, we started the estimation from the OLS values. The resulting LP estimates remain close to the starting values, as observed in previous studies.

Columns 6-7 report estimates using the GNR estimator and the modification to accommodate monopolistic competition (GNR-MC). Both handle transmission bias well in DGP1. The difficulties for these methods arise when monopolistic competition is introduced in DGP2. GNR show that under monopolistic competition their baseline 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, using a market-share-weighted average of deflated revenues as an aggregate demand shifter to construct the markup estimate (following De Loecker (2011)). This extension moves the estimates in the right direction but the variation in the aggregate demand shifter (which comes from variation in household income

[40] Given that we have omitted labor in our simulation, the LP estimator is similar to the Ackerberg et al. (2015) approach, which is the specific focus of the Rovigatti and Mollisi (2018) and Norris-Keiller et al. (2024) criticisms. An advantage of our TSIV approach is that it is not sensitive to starting values in this way.

in our context) is not sufficient. Greater variation in aggregate household income would improve the markup estimate in the simulation, but in real data it will be difficult to know whether the GNR-MC correction is adequate. Although in principle GNR and GNR-MC are not able to accommodate firm fixed effects,[41] the variance in the fixed effects here 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.[42]

Column 8 reports our TSIV estimates, using the second lags of materials and capital as internal instruments, and the log change of the international component of the materials price as an external instrument. Our estimates are robust to the inclusion of imperfect competition, firm effects, and input-quality differences. Our method yields tight confidence intervals, centered very close to the true values. Our approach compares well to the other methods in this admittedly artificial environment. The simulation leads us to have confidence that our method is a useful alternative to existing approaches in settings where valid external instruments are available, quantity information is observed, and it is reasonable to believe that our timing assumptions hold.

7 Conclusion

We have presented a new approach to improving production-function estimates by incorporating output and input quantity data and external instruments for materials and labor choices. To address the lack of an observable mapping between particular inputs and outputs, we have aggregated from the firm-product to firm level using CES aggregators. This in turn has made clear how quantity-based estimates of output elasticities are likely to be biased by quality and variety differences. To address the lack of a firm-level instrument for capital choices, we have developed a two-step instrumental variables (TSIV) approach in which we supplement the external instruments with internal instruments — lagged levels in the first-step difference equation and lagged differences in the second-step levels equation, in the spirit of System GMM. Our TSIV approach has the advantage that the estimates of the materials and labor coefficients are robust to mis-specification of the second step, which requires stronger assumptions. The difference in point estimates between our approach and those of other standard methods, although generally not statistically significant, carry potentially important economic implications. A simple Monte Carlo exercise suggests that our estimator performs better than standard estimators in an artificial setting with imperfect competition, firm fixed effects, and input quality differences.

Our approach requires richer data than is typically available, but the data frontier is expanding

[41] See Section VII.C.3 of GNR for a discussion of an extension to accommodate firm fixed effects.

[42] In 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.

rapidly, and we expect that researchers will increasingly have access to quantity information and the microdata required to construct credible external instruments at the firm level. Data on physical quantities of both inputs and outputs are already available in several countries, including Spain, Portugal, the US, Chile, Ecuador, China, and India. Future work may not use the same external instruments that we do here, but if researchers are encouraged to search for other external instruments, an important goal of this paper will have been accomplished.

A natural question is how to address quality and variety biases if one does not have external instruments that are credibly uncorrelated with unobserved quality and variety of inputs and outputs. One potential way forward would be to construct proxies for the quality and variety terms. 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 construct 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.[43]

Our main objective in this paper has been to improve estimates of output elasticities, which can be inserted directly into formulas for markups based on Hall (1988) and De Loecker and Warzynski (2012), as discussed above. But it is also natural to ask how such estimates can be used to improve estimates of productivity. This task is not straightforward. Two types of TFP measures have received particular attention: TFPQ, which uses physical output and input quantities, and TFPR, which uses (sector-level-deflated) 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 excludes idiosyncratic price variation, it may still reflect quality or variety differences on the input or output side. By contrast, TFPR does not misinterpret quality and variety differences as technical efficiency, but it may reflect 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 favor using TFPR, while keeping in mind that in general it will reflect both technical efficiency and price variation. Appendix Table A10 shows that TFPR constructed using our output-elasticity estimates is correlated with TFPR calculated using other methods, 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. One way forward would be to use our output-elasticity measures to improve estimates of markups and then use the markup estimates to correct TFPR along the lines of Garcia-Marin and Voigtländer (2019). We leave this task for future

[43]Note 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).

work.

References

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.

Ackerberg, Daniel A. and Jan De Loecker, "Production Function Identification Under Imperfect Competition," 2024. Unpub. paper.

_ , 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," *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.

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.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The Fall of the Labor Share and the Rise of Superstar Firms," *Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Bartelsman, Eric J. and Mark Doms, "Understanding Productivity: Lessons from Longitudinal Microdata," *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 When Firms Sort into Markets: The Case of Exporters," *Journal of Political Economy*, 2024, *132* (4), 1162–1208.

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 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.

Caselli, Mauro, Arpita Chatterjee, and Shengyu Li, "Productivity and Quality of Multi-product Firms," 2023. Unpub. paper.

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," in "Handbook of Industrial Organization," Vol. 4, Elsevier, 2021, pp. 141–223.

_ 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.

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.

_ and _ , "Reexamining the De Loecker & Warzynski (2012) Method for Estimating Markups," 2023. Unpub. paper, University of Pennsylvania.

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 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.

_ , _ , and Nicolas Urdaneta, "The Size and Life-Cycle Growth of Plants: The Role of Productivity, Demand, and Wedges," *Review of Economic Studies*, 02 2023, *91* (1), 259–300.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *American Economic Review*, 1994, *84* (1), 157 – 177.

Forlani, Emanuele, Ralf Martin, Giordano Mion, and Mirabelle Muûls, "Unraveling Firms: Demand, Productivity and Markups Heterogeneity," *Economic Journal*, 04 2023, *133* (654), 2251–2302.

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.

Hahn, Nadine, "Product differentiation and quality in production function estimation," 2024. ZEW Discussion Paper no. 24-049.

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.

Karabarbounis, Loukas and Brent Neiman, "The Global Decline of the Labor Share," *Quarterly Journal of Economics*, 10 2014, *129* (1), 61–103.

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.

Li, Jing, Shengyu Li, and Hongsong Zhang, "Output Quality, Productivity, and Demand: Evidence from the Chinese Steel Industry," 2023. Unpub. paper.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications," *Quarterly Journal of Economics*, 2020, *135* (2), 561–644.

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.

Nerlove, Marc, "Returns to Scale in Electricity Supply," in Carl F. Christ, ed., *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, Stanford University Press, 1963.

Norris-Keiller, Agnes, Áureo de Paula, and John Van Reenen, "Production Function Estimation Using Subjective Expectations Data," 2024. NBER working paper no. 32725.

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, Xinzheng 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," 2023. NBER working paper no. 31233.

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.

Restuccia, Diego and Richard Rogerson, "Misallocation and Productivity," *Review of Economic Dynamics*, 2013, *16* (1), 1–10.

Roodman, David, "How To Do xtabond2: An Introduction to Difference and System GMM in Stata," *Stata Journal*, 2009, *9* (1), 86–136.

_ , "A Note on the Theme of Too Many Instruments," *Oxford Bulletin of Economics and statistics*, 2009, *71* (1), 135–158.

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.

Sato, Kazuo, "The Ideal Log-Change Index Number," *Review of Economics and Statistics*, 1976, *58*, 223–228.

Shea, John, "The Input-Output Approach to Instrument Selection," *Journal of Business and Economic Statistics*, 1993, *11* (2), 145–155.

Shenoy, Ajay, "Estimating the Production Function under Input Market Frictions," *Review of Economics and Statistics*, 09 2021, *103* (4), 666–679.

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.

Syverson, Chad, "Market Structure and Productivity: A Concrete Example," *Journal of Political Economy*, 2004, *112* (6), 1181–1233.

_ , "What Determines Productivity?," *Journal of Economic Literature*, 2011, *49* (2), 326–65.

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*, 01 2023, *90* (6), 3315–3342.

Vartia, Yrjo, "Ideal Log-Change Index Numbers," *Scandinavian Journal of Statistics*, 1976, *3*, 121–126.

Verhoogen, Eric, "Firm-Level Upgrading in Developing Countries," *Journal of Economic Literature*, 2023, *61* (4), 1410–1464.

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***	0.675***	0.488***	0.485***
	(0.027)	(0.027)	(0.050)	(0.052)
log labor (ℓ_{it})	0.298***	0.296***	0.294***	0.288***
	(0.039)	(0.040)	(0.037)	(0.036)
log capital (k_{it})	0.087***	0.087***	0.010	0.010
	(0.019)	(0.019)	(0.019)	(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 ($\widetilde{y}_{it}^{SV}$)				
Sato-Vartia materials index ($\widetilde{m}_{it}^{SV}$)	0.469***	0.468***	0.434***	0.428***
	(0.084)	(0.085)	(0.052)	(0.053)
log labor (ℓ_{it})	0.357***	0.358***	0.283***	0.274***
	(0.109)	(0.110)	(0.046)	(0.045)
log capital (k_{it})	0.196***	0.195***	0.013	0.015
	(0.045)	(0.046)	(0.020)	(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

	$\triangle\widetilde{m}_{it}^{SV}$	$\triangle\ell_{it}$	$\triangle k_{it}$	$\triangle\widetilde{m}_{it}^{SV}$	$\triangle\ell_{it}$	$\triangle k_{it}$	$\triangle\widetilde{m}_{it}^{SV}$	$\triangle\ell_{it}$	$\triangle k_{it}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$\widetilde{m}_{it-2}^{SV}$	-0.017***	0.013***	0.027***				-0.018***	0.012***	0.026***
	(0.006)	(0.004)	(0.005)				(0.006)	(0.004)	(0.005)
ℓ_{it-2}	0.014	-0.030***	0.044***				0.013	-0.030***	0.044***
	(0.009)	(0.007)	(0.010)				(0.009)	(0.007)	(0.010)
k_{it-2}	0.009	0.009**	-0.048***	0.001	0.001	-0.010***	0.007	0.007*	-0.050***
	(0.006)	(0.004)	(0.007)	(0.003)	(0.002)	(0.003)	(0.006)	(0.004)	(0.007)
$\triangle$ pred. import price index ($\triangle\widetilde{w}_{it}^{imp}$)				-0.248**	-0.039	0.072	-0.255***	-0.045	0.118
				(0.098)	(0.064)	(0.103)	(0.098)	(0.063)	(0.102)
$\triangle$ min. wage x "bite" ($\triangle z_{it}$)				-1.366	-2.077***	-2.398***	-1.492	-2.062***	-1.991***
				(1.046)	(0.545)	(0.623)	(1.049)	(0.549)	(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.: $\triangle$ log output index ($\triangle\widetilde{y}_{it}^{SV}$)			
	internal instruments only	external instruments + k_{it-2}	internal & external instruments	internal & external instruments (LIML)
	(1)	(2)	(3)	(4)
$\triangle$ log materials index ($\triangle\widetilde{m}_{it}^{SV}$)	0.520	0.240	0.449**	0.450**
	(0.487)	(0.528)	(0.192)	(0.195)
$\triangle$ log labor ($\triangle\ell_{it}$)	0.485	0.688	0.471***	0.472***
	(0.394)	(0.556)	(0.176)	(0.177)
$\triangle$ log capital ($\triangle k_{it}$)	-0.148	-0.166	-0.153	-0.154
	(0.196)	(0.243)	(0.133)	(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)
$\triangle$ log capital, lagged ($\triangle k_{it-1}$)	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.: $\widetilde{y}_{it}^{SV} - \widehat{\beta}_m \widetilde{m}_{it}^{SV} - \widehat{\beta}_\ell \ell_{it}$	
	IV (1)	OLS (2)
log capital k_{it}	0.114	0.151***
	[0.200]	(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 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 3.6 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 ($\triangle\widetilde{y}_{it}^{SV}$)		
	(1)	(2)	(3)
$\widetilde{y}_{it-1}^{SV}$	1.008***	0.991***	0.956***
	(0.038)	(0.036)	(0.048)
$\widetilde{m}_{it}^{SV}$	0.589***	0.659***	0.455***
	(0.125)	(0.097)	(0.063)
$\widetilde{m}_{it-1}^{SV}$	-0.596***	-0.662***	-0.401***
	(0.124)	(0.105)	(0.069)
log labor$_{it}$ (ℓ_{it})	0.310	0.203	0.292***
	(0.257)	(0.260)	(0.110)
log labor$_{it-1}$ (ℓ_{it-1})	-0.315	-0.216	-0.315***
	(0.238)	(0.235)	(0.092)
log capital$_{it-1}$ (k_{it})	0.070	0.139	0.106*
	(0.094)	(0.085)	(0.054)
log capital$_{it-1}$ (k_{it-1})	-0.070	-0.112	-0.095**
	(0.089)	(0.076)	(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 (2009a), with initial weighting matrix defined in Doornik et al. (2012) and finite-sample correction from Windmeijer (2005). The Stata command is xtabond2 (Roodman, 2009a), 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. 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 6 for further details. Robust standard errors in parentheses. *10% level, **5% level, ***1% level.

Table 6. Comparison to Other Methods

	OP (1)	LP (2)	GNR (3)	GNR-MC (4)	TSIV (5)
$\widehat{\beta}_m$ (materials)	0.669***	0.639***	0.406***	0.560***	0.449**
	(0.028)	(0.049)	(0.009)	(0.073)	(0.195)
$\widehat{\beta}_\ell$ (labor)	0.254***	0.291***	0.513***	0.597***	0.471***
	(0.060)	(0.045)	(0.037)	(0.085)	(0.177)
$\widehat{\beta}_k$ (capital)	0.134***	0.108***	0.138***	0.272**	0.114
	(0.044)	(0.041)	(0.027)	(0.116)	(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 6 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	FD	SysGMM	OP	LP	GNR	GNR-MC	TSIV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. DGP1: Perfect Competition								
$\widehat{\beta}_m$ (materials)	0.843	0.914	0.810	0.843	0.847	0.650	0.650	0.653
	(0.003)	(0.006)	(0.083)	(0.006)	(0.010)	(0.000)	(0.000)	(0.017)
$\widehat{\beta}_k$ (capital)	0.112	0.061	0.137	0.112	0.110	0.250	0.250	0.248
	(0.002)	(0.004)	(0.063)	(0.004)	(0.007)	(0.000)	(0.000)	(0.012)
B. DGP2: Imperfect Competition								
$\widehat{\beta}_m$ (materials)	0.735	1.016	0.713	0.739	0.748	0.550	0.596	0.650
	(0.004)	(0.008)	(0.065)	(0.005)	(0.014)	(0.000)	(0.059)	(0.021)
$\widehat{\beta}_k$ (capital)	0.209	0.078	0.220	0.208	0.203	0.298	0.231	0.249
	(0.002)	(0.004)	(0.040)	(0.003)	(0.007)	(0.000)	(0.023)	(0.010)
C. DGP3: Imperfect Competition, Firm Effects								
$\widehat{\beta}_m$ (materials)	0.819	1.016	0.721	0.826	0.834	0.550	0.576	0.648
	(0.003)	(0.007)	(0.064)	(0.004)	(0.015)	(0.000)	(0.043)	(0.020)
$\widehat{\beta}_k$ (capital)	0.170	0.078	0.216	0.167	0.163	0.300	0.225	0.251
	(0.002)	(0.004)	(0.040)	(0.003)	(0.007)	(0.000)	(0.016)	(0.010)
D. DGP4: Imperfect Competition, Firm Effects, Input-Quality Differences								
$\widehat{\beta}_m$ (materials)	0.854	1.075	0.773	0.861	0.869	0.550	0.591	0.648
	(0.003)	(0.007)	(0.061)	(0.005)	(0.013)	(0.000)	(0.047)	(0.021)
$\widehat{\beta}_k$ (capital)	0.153	0.051	0.187	0.151	0.146	0.300	0.231	0.251
	(0.002)	(0.004)	(0.037)	(0.003)	(0.006)	(0.000)	(0.018)	(0.010)

Notes: Table presents Monte Carlo output-elasticity estimates for four data-generating processes (DGPs), all with serially independent productivity shocks. See Section 6.2 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 150,000 observations each (except for OP, where we drop observations with zero investment).