DISCUSSION PAPER SERIES

DP17079
(v. 2)

Production function estimation using observed price changes

Xulia González, Saul Lach and Daniel Miles

INDUSTRIAL ORGANIZATION

ISSN 0265-8003

Production function estimation using observed price changes

Xulia González, Saul Lach and Daniel Miles

Discussion Paper DP17079
First Published 02 March 2022
This Revision 25 April 2022

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Production function estimation using observed price changes

Abstract

The estimation of production functions with firm-level data is hampered by the non-observability of real quantities (output and inputs). An increasing number of papers use reported price changes to construct firm-specific price indices which are then used to deflate nominal quantities. We argue that this procedure does not address the fundamental problem of the non-observability of prices because it introduces an initial condition given by the price level in the base year chosen to construct the price index. Because this price level is unobserved, this procedure generates an omitted variable bias (OVB) whose importance can be gauged by the non-robustness of the estimated production function parameters to the choice of base year. The OVB disappears when firm-specific fixed effects are incorporated in the estimation of the production function. We illustrate using simulations and a panel of Spanish manufacturing firms.

JEL Classification: D22, D24, L10, L60

Keywords: Production function estimation, Productivity, unobserved prices, omitted variable bias

Xulia González - xgzlez@uvigo.es
University of Vigo

Saul Lach - saul.lach@mail.huji.ac.il
The Hebrew University of Jerusalem and CEPR

Daniel Miles - dmiles@uvigo.es
University of Vigo

Acknowledgements
We thank Salvador Navarro for providing us with the basic computer code for implementing the Ghandi, Navarro and Rivers (2020) estimator. We also thank Omar Licandro, Andrea Pozzi, Mark Roberts and Fabiano Schivardi for useful comments and suggestions. The authors gratefully acknowledge funding from the Spanish Ministry of Science, Innovation and Universities and FEDER through Grant no. RTI-2018-099403-B-I00. The authors are solely responsible for any errors.

Production function estimation using observed price changes *

Xulia González† Saul Lach‡ Daniel Miles§

March 24, 2022

Abstract

The estimation of production functions with firm-level data is hampered by the non-observability of real quantities (output and inputs). An increasing number of papers use reported price changes to construct firm-specific price indices which are then used to deflate nominal quantities. We argue that this procedure does not address the fundamental problem of the non-observability of prices because it introduces an initial condition given by the price level in the base year chosen to construct the price index. Because this price level is unobserved, this procedure generates an omitted variable bias (OVB) whose importance can be gauged by the non-robustness of the estimated production function parameters to the choice of base year. The OVB disappears when firm-specific fixed effects are incorporated in the estimation of the production function. We illustrate using simulations and a panel of Spanish manufacturing firms.

Keywords: production function estimation, productivity, unobserved prices, omitted variable bias.

JEL Codes: D22, D24, L10, L60

* We thank Salvador Navarro for providing us with the code of the Ghandi, Navarro and Rivers (2020) estimator. We also thank Omar Licandro, Andrea Pozzi, Mark Roberts and Fabiano Schivardi for useful comments and suggestions. The authors gratefully acknowledge funding from the Spanish Ministry of Science, Innovation and Universities and FEDER through Grant no. RTI-2018-099403-B-I00. The authors are solely responsible for any errors.

† Universidade de Vigo, Spain; E-mail: xgzlez@uvigo.es; http://webs.uvigo.es/xgzlez

‡ The Hebrew University of Jerusalem and CEPR. E-mail: Saul.Lach@mail.huji.ac.il; http://saullach.weebly.com

§ Universidade de Vigo, Spain; E-mail: dmiles@uvigo.es; http://webs.uvigo.es/dmiles

1 Introduction

The estimation of production functions using firm level data plays a key role in many empirical analyses. Meaningful policy parameters, policy evaluations and aggregate productivity growth require knowledge of the production function. Its estimation, however, raises a set of empirical issues that have no straightforward solution (Bond et al., 2021; Pakes, 2021; Syverson, 2011).

In this paper we focus on one such problem: the lack of data on physical quantities.[1] In practice, output and inputs are measured by nominal revenues and expenditures deflated by a price index. When firm-specific prices are not observed, an industry-wide price index is often used to deflate the nominal values. As shown by Klette and Griliches (1996), this approach generates an omitted variable problem defined by the difference between the industry price index and the firm's specific price index (de Loecker, 2011).

To avoid this problem, recent research started using additional price information reported in firm-level surveys to construct firm-specific price indices. A common practice is to use reported annual changes in prices, and a recursive formula, to generate a firm-specific price level. This is feasible because an increasing number of surveys are now reporting firm level price changes.[2] Even when the price levels themselves are reported, such as in the Colombian Encuesta Anual de Manufacturas (EAM), researchers ignore this information and chose, instead, to construct their price indices based on price changes.[3]

In this paper we show that this approach – deflating nominal values by a firm-specific price index constructed from price changes – leads to an omitted variable problem similar, in spirit, to that found by Klette and Griliches (1996). The recursive formula used to construct the firm-specific price index using reported price changes depends on an initial condition given by the unobserved firm-specific

[1] Another well-known problem arises when the firm's optimal choice of inputs depends on its productivity, making these inputs endogenous in the production function. Structural estimation methods are increasingly used to address this problem, starting with Olley and Pakes (1996) and following with Levinsohn and Petrin (2003), Ackerberg et al. (2015) and others. These authors adopt a control function or proxy-variable approach to estimating the production function. Recently, Gandhi, et al. (2020) offer a nonparametric procedure that identifies an unknown production function using the proxy-variable approach.

[2] For example, the Spanish Encuesta Sobre Estrategias Empresariales (ESEE) survey, the Bank of Italy's Survey on Italian Manufacturing Firms (INVIND), and the French Manufacturing Survey, among others, report firm specific annual price changes.

[3] See Eslava et al (2004). This paper has been quite influential since its methodology for constructing firm-specific price indeces has been widely adopted (Smeets and Warzynski, 2013; Carlsson et al., 2021).

price level in the chosen base year. Because firms' output prices are likely to be correlated over time, the price level in the base year is likely to be correlated with input choices in all periods. This generates an additional source of endogeneity, besides the one generated by unobserved productivity, that can bias the estimation of production functions.

The usual solution to the missing base year price level is to ignore it. This is what happens when the firm-specific base year price is normalized to a common price across firms, usually set to unity (or zero in logarithms). Notice that this is equivalent to treating the price in the base year as an unobserved component of the error term that is uncorrelated with input choices. As mentioned above, ignoring the endogenous variation in the base year price level is likely to bias the estimation of production function parameters. Another reason why such a normalization is not warranted is that, when the year chosen as base year is the same for all firms, normalizing the price to unity is not internally consistent since it forces price to be homogeneous across firms in the base year, despite the existence of price heterogeneity in other years. In fact, normalization of prices runs against the documented heterogeneity across firms in a variety of dimensions (Bartelsman and Doms, 2000)

A contribution of the paper is to point out the existence of an endogeneity problem resulting from the use of firm-specific price indices, based on reported price changes, to deflate nominal revenues and input expenditures. But the paper also offers a simple way of solving this problem using recent econometric methods based on the idea that the omitted price level is constant over time but differs across firms and can therefore be treated as a firm specific fixed-effect. Indeed, incorporating fixed effects in the estimation of the production function removes the omitted variable bias.

A stark manifestation of the Omitted Variable Bias (OVB) introduced by the unobserved base year price is that the estimated production function parameters change when one changes the base year used to construct the price index. In general, one should not expect the choice of base year to affect the estimated production function. But this is not the case when the base year price is unobserved because the magnitude of the omitted variable bias depends on the choice of base year. Using simulations and a data set of Spanish manufacturing firms we report significant changes in the estimated production function elasticities when different base years are chosen. These large changes in estimated parameters reflect the importance of the omitted base year price level. This is an awkward situation because the choice of base year is at the researcher's discretion.

The main practical recommendation of this paper is that when using information on price changes to deflate nominal quantities, one needs to incorporate fixed effects in the estimation of the production function to account for the unobserved base year prices. Although traditional panel data estimation methods did account for fixed effects, these has been largely ignored in the more recent control/proxy-variable estimation approach. However, as shown by Ghandi et al. (2020), under additional assumptions, fixed effects can be re-introduced into this estimation approach. Furthermore, we recommend showing the robustness of the estimated objects of interest (e.g., elasticities, productivity, etc.) to the choice of base year.

The paper is organized as follows. In Sections 2 and 3 we show analytically and using simulated data that when the endogeneity of the base year output price is ignored (either by normalizing it or by treating it as an uncorrelated error component), estimates of production function parameters based on firm-level price changes are biased. Introducing fixed-effects in the estimation solves this problem. Extensions are given in Section 4, while in Section 5 we illustrate the problem and solution using a popular data set of Spanish manufacturing firms. In Section 6 we adapt the methodology in Gandhi et al. (2020) to estimate revenue functions with unobserved productivity and fixed effects, when firm prices are partially observed (i.e., only their changes over time are observed). In Section 7, we apply the Ghandi el al. estimator to the aforementioned dataset to illustrate the non-robustness of the estimated results to the choice of base year. In this Section, we also compute aggregate productivity and examine how it changes with the choice of base year. We also compare estimated aggregate productivity across estimators that account for or ignore the unobserved price in the base year. The impact of these estimators on productivity dispersion is also discussed. Summary and conclusions close the paper.

2 The bias from the unobserved base year price

2.1 The base year price as a fixed effect

The production function of firm i is $Y_{it} = F(k_{it}, l_{it}, m_{it})e^{\omega_{it}+\eta_{it}}$ and in logarithms, denoted by lowercase letters,

$$y_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} \tag{1}$$

where $i = 1, \ldots, N_t$ and N_t is the number of firms in year t. ω_{it} is productivity, k_{it}, l_{it}, m_{it} are measures of capital, labor and materials (in logs), and η_{it} is an I.I.D. zero mean shock to production.

To estimate equation (1) we need to address the endogeneity problem introduced by the unobserved (to the researcher) productivity term ω_{it}, as well as measurement issues that arise because firm-level datasets usually do not report the specific variables that appear in (1). In this Section we discuss the measurement problems associated with the non-observability of output.[4]

Output y_{it} is usually not observed because most firm-level datasets report revenues instead of physical output. When we substitute revenue for output the production function equation becomes

$$r_{it} = y_{it} + p_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} + p_{it} \tag{2}$$

where r_{it} are log revenues and p_{it} is log output price.

Clearly, the non-observability of output would not be an issue if firm's output price levels p_{it} were observed. However, price levels are not usually reported so that p_{it} is unobserved. Because price levels are likely to be correlated with input choices, ignoring the p_{it} in equation (2) generates an omitted variable bias (OVB). Klette and Griliches (1996) show that using an industry-wide price index instead of p_{it} to deflate revenues does not correct for this bias (see also De Loecker, 2011 and 2021).

Recently, an increasing number of firm-level surveys report output price changes, i.e., $\Delta P_{it} = \frac{P_{it} - P_{it-1}}{P_{it-1}}$. Researchers have used these reported price changes to construct a firm-specific price index, thereby hoping to overcome the problem of not observing price levels.[5] In this paper we argue that doing this does not solve the omitted variable bias. In fact, when using price changes to construct a firm-specific price index, an omitted variable is introduced into the estimated equation given by the unobserved price level in the base year chosen to construct the index.

The idea behind using price changes is that the unobserved price P_{it} can be recovered from these

[4] In Section 4.1 we analyze the lack of data on physical inputs, while in Section 6 we address the endogeneity problem associated with unobserved productivity.

[5] Examples of surveys where price changes are reported appear in footnote 2. Papers using price changes to construct firm-specific price index are, for example, Eslava et al., 2004; Mairesse and Jaumandreu, 2005; Smeets and Warzynski, 2013; Dolado et al. 2016; Jaumandreu and Lin, 2018; Carlsson et al., 2021.

changes using the recursive formula

$$P_{it} = P_{it-1}(1 + \Delta P_{it}) = P_{ib} \prod_{s=b+1}^{t} (1 + \Delta P_{is}) = P_{ib} \prod_{s=b+1}^{t} \frac{P_{is}}{P_{is-1}}$$

for $t > b$.

Notice that the initial condition in this formula is the *price level in the base year* P_{ib}. The base year b is chosen by the researcher and can be any of the years observed in the sample. For clarity of exposition we focus on the casse $b < t$. [6] Of course, because price levels are not observed, P_{ib}, is not observed.

Log prices are then

$$p_{it} = \begin{cases} p_{ib} + \sum_{s=b+1}^{t} \Delta p_{is} & t > b \\ p_{ib} & t = b \end{cases} \tag{3}$$

where $\Delta p_{is} = \ln\left(\frac{P_{is}}{P_{is-1}}\right) = p_{is} - p_{is-1}$.

Substituting (3) for p_{it} in equation (2) results in

$$r_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} + p_{ib} + \sum_{s=b+1}^{t} (\Delta p_{is})$$

for $t \geq b + 1$, delivering the estimable equation

$$r_{it}^{*} \equiv r_{it} - \sum_{s=b+1}^{t} (\Delta p_{is}) = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} + p_{ib}. \tag{4}$$

We remark that while price changes $\sum_{s=b+1}^{t}(\Delta p_{is})$ and therefore r_{it}^{*} are observed, the price level

[6] Two remarks are in order. First, we do not specify which calendar year is chosen as base year by the researcher. The base year can vary across firms and can even be a year outside the period in which the firm is observed in the sample. In practice, researchers usually select a common base year for all firms. In Section 4.2, we explicitly consider the case where firms enter the sample at different points in time and argue against this common procedure, while proposing a natural alternative.

Second, if $b > t$ the recursive formula changes. In general, for any t and b the formula is,

$$p_{it} = \begin{cases} p_{ib} + \sum_{s=b+1}^{t} \Delta p_{is} & t \geq b + 1 \\ p_{ib} & t = b \\ p_{ib} - \sum_{s=t+1}^{b} \Delta p_{is} & t \leq b - 1 \end{cases}$$

in the base year p_{ib} is not. The observed r_{it}^* are "partially deflated" revenues because the base year price level is not included in the deflation. Since $r_{it}^* = y_{it} + p_{ib}$, partially deflated revenues provide an error-ridden measurement of physical output. Thus, p_{ib} essentially represents a measurement error which is constant over time but may vary across firms.

Most empirical papers assume away the unobserved price p_{ib} by setting $P_{ib} = 1$, so that $p_{ib} = 0$ for all i, i.e., all firms have the same price in the base year.[7] By ignoring the variation in p_{ib} across firms, this normalization is essentially equivalent to assuming that p_{ib} can be treated as a random error component in (4) uncorrelated with input choices. Thus, the seemingly trivial normalization of the base year price level entails a strong behavioral assumption. Any reasonable model of firm behavior will imply a correlation between input choices and output price. Furthermore, persistence over time in output prices implies that p_{ib} is likely to be correlated with input choices in *all* periods. If this endogenous variation across firms in p_{ib} is ignored, i.e., either by assuming $p_{ib} = 0$ for all i or by actually treating it as an uncorrelated error component, it will bias the estimation of the production function.

This bias does not depend on the way the firm decides upon prices. The bias exists even if prices are exogenous to the firm (unrelated to productivity) as long as input demands respond to output price, output prices vary across firms and are persistent over time, for which there is ample empirical evidence.[8]

Thus, using reported price changes and estimating (4) while ignoring p_{ib} is likely to result in biased estimators of the production function. This endogeneity problem results from the lack of full information on prices: observing firm-level prices *changes* is not enough to address the lack of

---

[7] Setting the price level to unity is just a convenient normalization: any other constant distinct from zero will be absorbed into the constant term of the production function. We emphasize that a unity price in the base year is an assumption on the firm output price *level*, i.e., $P_{ib} = 1$, not to be confused with the (unobserved) firm-specific price index being one in the base year which is so by definition since $\frac{P_{it}}{P_{ib}} \equiv 1$ when $t = b$.

[8] The findings in the literature suggest that firm output price levels are persistent over time and/or are a function of firm's productivity which is also persistent (e.g., Roberts and Supina, 2000; Foster et al., 2008).

information on price *levels*.[9,10]

The main message of this paper is that we should treat the omitted regressor p_{ib} in equation (4) as an unobserved firm-level fixed effect – potentially correlated with the input regressors – and apply suitable estimation methods.[11]

2.2 Choice of base year

It is intuitive to require that the estimands of the production function remain invariant to the choice of base year, which is at the researcher's discretion. This is so because

$$r_{it} = y_{it} + p_{it} = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} + p_{it}$$

and changing the base year does not affect this relationship.[12] Put differently, the choice of base year does not affect the identification of the production function.

This, however, is not necessarily the case when p_{ib} is unobserved and normalized to zero, i.e.,

---

9 When price changes are not observed it is common to use an industry-level (or any other grouping) output price deflator, under the assumption that prices change at the same rate for all firms in the same industry. Doing this does not solve the problem of the unobserved p_{ib}. Let the industry-level price index be P_t/P_b where P_t is the unobserved industry-level price in year t. Deflated revenues are then

$$R_{it}^d = \frac{P_{it}Y_{it}}{P_t/P_b} = P_{ib}Y_{it}\frac{P_{it}/P_{ib}}{P_t/P_b}$$

and assuming that prices change at the same rate for all firms implies $\frac{P_{it}/P_{ib}}{P_t/P_b} = 1$, so that $R_{it}^d = P_{ib}Y_{it}$ or

$$r_{it}^d = f(k_{it}, l_{it}, m_{it}) + \omega_{it} + \eta_{it} + p_{ib}$$

as in (4). That is, the same unobserved fixed effect p_{ib} is present when applying an industry-level price deflator.

10 When capital is measured using a permanent inventory method it also depends on an unobserved initial capital level. The impact of this initial condition, however, diminishes over time because of capital depreciation. In our case, the impact of the initial price level does not vanish over time because of the presence of a unit root in (3). We thank Fabiano Schivardi and Andrea Pozzi for this observation.

11 In practice, using fixed effects (within) estimators in the estimation of production functions needs to deal with the small within-firm variation in the capital measure which results in imprecise estimators of the capital elasticity. One possibility, explored by Pozzi and Schivardi (2016), is to use capital utilization rates which fluctuate considerably over time. If capital utilization is not reported and fixed effects are not used, then one should delete period $t = b$ observations from the regression. Doing this will prevent any bias generated by the correlation between p_{ib} and inputs at time $t = b$ and may help in reducing the OVB, although is not likely to eliminate it.

12 The output price level can be written as $p_{it} = p_{ib} + p_{it} - p_{ib} = p_{ib} + \sum_{s=b+1}^{t} \Delta p_{is}$ so that changing the base year from $t = b$ to $t = b' > b$ is equivalent to adding and subtracting $p_{ib'}$, $p_{it} = p_{ib'} + p_{ib} - p_{ib'} + \sum_{s=b+1}^{t} \Delta p_{is} = p_{ib'} + \sum_{s=b'+1}^{t} \Delta p_{is}$, which does not affects p_{it}.

omitted in the estimation of equation (4). A change in the estimated equation, after the base year is changed, reflects differences in the magnitude of the OVB and is therefore indicative of a correlation between the unobserved base year price level and the regressors. Indeed, in our empirical application in Section 5, using a sample of Spanish manufacturing firms, we find measurable changes in the estimated production elasticities when the base year is changed (also in our simulation results in subsection 3.2.2).

Of course, using fixed effects to account for the unobserved p_{ib} deals with this problem. We therefore recommend (re-)introducing fixed effects into the estimation of production functions when observing only price changes.[13] We show how to do this using the GNR estimator in Section 6 and apply it to the Spanish data in Section 7.

Finally, an additional implication of the presence of p_{ib} in (4) is that, even if the base year price is uncorrelated with the input variables, the control approach procedure would recover an estimate of productivity that includes a demand component captured by p_{ib}. Thus, even if ignoring differences in p_{ib} across firms does not bias the estimation of the production function, there is still need to account for p_{ib} when the goal is to estimate productivity levels. We illustrate this in subsection 7.1.

3 An illustrative simple model

In this Section we present a simple model of firm behavior to illustrate how the unobserved base year price level can bias the production function estimators. The model in subsection 3.1 shows how prices in the base year can be correlated with input choices in other years. In subsection 3.2 we use the model to simulate data which will enable us to quantify the bias associated with the arbitrary normalization of the output price in the base year.

[13] Somewhat ironically, traditional panel data estimation methods do account for firm fixed effects and therefore are not subject to this OVB. However, the recent estimators based on a proxy-variable/control approach, initiated by Olley and Pakes (1996) and Levinsohn-Petrin (2003), do not typically allow for firm-level fixed effects in their estimation procedures (Ackerberg, 2021).

3.1 A simple model

The firm's production function is Cobb-Douglas with a single variable input L (labor),

$$Y_{it} = L_{it}^{\alpha} e^{\omega_{it}+\eta_{it}} \tag{5}$$

The firm observes the exogenous ω_{it}, but not η_{it}, when making its input decision. Furthermore, we assume firm i is a price-taker and that output price P_{it} may vary across firms and over time. This is a simplifying assumption and a more realistic model should let the firm choice prices given its observed productivity level and demand factors. Nevertheless, the simple model conveys the main message of the paper.[14]

The profit-maximizing amount of labor is, in logs,

$$l_{it} = \beta_0 - \frac{1}{(1-\alpha)} p_{Lit} + \frac{1}{1-\alpha} \left(\omega_{it} + p_{it}\right) \tag{6}$$

where $\beta_0 = \frac{1}{(1-\alpha)} \ln\left(\alpha E(e^{\eta_{it}})\right)$, $p_{it} = \ln(P_{it})$ and $p_{Lit} = \ln(P_{Lit})$ is log wages.

In this simple model there are three exogenous firm-specific variables – productivity ω_{it}, prices p_{it} and wages p_{Lit} – driving the endogenous variables: input L and output Y. When we estimate the production function using nominal revenues $r_{it} = y_{it} + p_{it}$ we are estimating a parametric version of (2), namely,

$$r_{it} = \alpha l_{it} + \omega_{it} + \eta_{it} + p_{it} \tag{7}$$

Labor is endogenous in (7) because it is correlated with productivity ω_{it} *and* with price p_{it}. Note that the correlation between labor and output price is due to the variation in price across firms, and not to market power per-se. Moreover, in this simple model, unobserved productivity and unobserved price are equally correlated with labor input and, therefore, are equally "responsible" for its endogeneity in the production function.

If observed, adding firm-level prices to the right hand side of the estimating equation or, equivalently, deflating nominal revenues by the firm's own prices, would solve this dimension of labor's endogeneity. When *only* price changes are observed we can only partially deflate revenues, leading

[14] Appendix A presents an illustrative example of a monopolistic competitive firm.

to

$$r_{it}^* = r_{it} - \sum_{s=b+1}^{t} \Delta p_{is} = \alpha l_{it} + \omega_{it} + \eta_{it} + p_{ib} \tag{8}$$

corresponding to the general case (4).

When the price level in the base year p_{ib} is not observed the question is whether labor in year t, l_{it}, and p_{ib} are correlated. Clearly, there is a correlation when $t = b$ because labor is correlated with the contemporaneous price level. Because l_{it} is a function of p_{it}, the correlation in other time periods depends on the serial correlation in prices. As long as there is some serial correlation, it is likely that l_{it} will be correlated with p_{ib} thereby making labor endogenous in the production function. This source of endogeneity is in addition to the classical endogeneity arising form omitting productivity when estimating the production function.

To get a quantitative understanding of the bias caused by the omitted variation in the base year price, we simulate data from this simple model and use them to estimate the production function parameters using (8).

3.2 Simulation results

We simulate data based on the model in subsection 3.1 for 500 firms over 100 periods, but we use only the last 10 periods in estimation. Thus, we have a panel of 5,000 (500×10) observations. In most simulations we assume that the exogenous output price, wages and productivity processes are highly persistent (AR1 coefficient set to 0.8) and independent (details are in Appendix B). The labor parameter is set to $\alpha = 0.35$.

3.2.1 The bias from ignoring the base year price level

In this subsection we analyze the bias that arises when ignoring the unobserved base year price level in estimation. We use a simple OLS estimator, with and without firm fixed effects, to estimate variants of (8). Table 1 shows the average of the labor coefficient estimates – and its standard deviation – over 1000 repetitions of the sample data.

In the first column we report the results for the ideal case where firm-specific prices p_{it} and productivity ω_{it} are observed. In this case we are essentially estimating equation (7) and the OLS es-

Table 1: Estimates of (fully and partly) deflated revenue equation

	(1) p and ω observed	(2) p not observed	(3) p_b not observed	(4) p_b not observed	(5) p_b not observed serially uncorrelated prices	(6) p_b and ω not observed	(7) p_b and ω not observed
Dep. variable	$r - p = y$	r	r^*	r^*	r^*	r^*	r^*
Log labor	0.35	0.67	0.50	0.35	0.35	0.66	0.57
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Productivity	1.00	0.5	0.78	1.00	1.00		
	(0.01)	(0.03)	(0.04)	(0.02)	(0.03)		
Firm FE	NO	NO	NO	YES	NO	NO	YES

Entries are averages over 1000 repetitions of the estimated parameters in each simulation. The number of firms and periods in each simulated sample are 500 and 10, respectively. Labor elasticity is 0.35. The autoregressive parameters for wages and productivity are set to 0.8 in all columns.
The autoregressive parameter for prices is 0.8 in all columns except in column (5) where it is zero.
The dependent variable in column (1) is fully deflated revenues ($r - p = y$), in column (2) it is nominal revenues, and in columns (3)-(7) it is partially deflated revenues ($r*$).
The sample in column (5) does not include the base year t = b.
Standard errors in parentheses are standard deviations of estimated parameters across the 1000 repetitions.

timator of α is unbiased. This is reflected in the close match between the average of the estimated parameters and their true value.

In order to focus solely on the effect of omitting the output price from the regression, we will continue to include the firm-specific productivity ω_{it} as a regressor. Thus, in column (2), we omit only the firm-specific price p_{it}. As shown by Klette and Griliches (1996), the OLS estimator of equation (7) is biased when prices are omitted and this is manifested in an upward bias of about 91 percent, induced by the positive correlation between labor demand and output price (see (6)).

In column (3) we assume that we observe the change in output price but not the price level in the base year $b = 1$. As explained in Section 2, partially deflating revenues by the cumulative sum of the log price changes from $t = 2$ until t still leaves the price in the base year as an unobserved component of the error term. Because the price level at t is correlated with the firm's labor choice at t and the price process is persistent over time, we can expect the unobserved base year price level to be correlated with labor in every year t. This, in turn, biases the OLS estimator of equation (8) as clearly seen in column (3) where the bias is about 43 percent.

In column (4) we add firm-level fixed effect to the OLS regression to deal with the unobserved base year price level, as explained in Section 2. With fixed effects, the omitted variable bias is indeed eliminated and the labor elasticity is estimated consistently.

Column (5) replicates column (3) using *serially uncorrelated* output prices (but the same serially correlated wages and productivity). As argued in Section 2, without persistence in the output price process, p_{ib} is not correlated with p_{it} and therefore also not correlated with labor input l_{it} in year t. Thus, in this case, the unobserved p_{ib} can be treated as a random effect.[15]

Finally, both the price in the base year p_{ib} *and* productivity ω_{it} were omitted in columns (6) and (7), without and with fixed effects, respectively. In both cases the bias persists because productivity is still omitted from the regression, although including fixed effects to control for p_{ib} reduces the magnitude of the bias. In Section 6 we show how to consistently estimate the production function in this case.

3.2.2 Changing the base year

As mentioned in Section 2, the estimates of the production function parameters should remain invariant to the choice of base year. If we find that they change when the base year is changed, this is indicative that the price level in the base year, p_{ib}, is correlated with the regressor l_{it}. We therefore reestimate the production function (8) without fixed effects using a *different base year* and check for differences in the estimated parameters. Note that we compare the same estimator using the same sample data except for the partially deflated revenues which change due to the change in base year.[16]

For each simulated sample we estimate (8) twice: once when the base year is $b = 1$ and another one when the base year is $b' = 5$ (recall that there are 10 periods).[17] We do this under several serial correlation scenarios for the price process ranging from highly persistent output prices (AR1 coefficient set to 0.8 as in subsection 3.2.1) to serially uncorrelated prices. The top panel of Table 2 shows statistics on the ratio of the estimated labor elasticity using base year $b' = 5$ to the estimate

[15] The sample in column (5) does not include the base year, i.e., we use periods $t \neq b$, in order to avoid the correlation between labor demand in period $t = b$ and the unobserved base year price p_{ib}.

[16] We remark that this procedure is not a test of the null hypothesis of lack of an OVB. It only detects changes in the magnitude of the OVB. That is, the finding of no impact of a change in the base year is consistent with an OVB that is constant over time, e.g., when prices are constant over time so that $p_{ib} = p_{ib'}$. One could perform a type of Hausman test to test for the lack of an OVB but what we propose is much simpler.

[17] Years $b = 1$ and $b' = 5$ are not included in the estimation. Of course, one could do this for any year in the sample period.

Table 2: Ratio of estimated labor elasticity (base year 5 to base year 1)

	Output price AR1 coefficient ($\rho = 0$)			
	0.8	0.5	0.25	0
			Without Fixed Effects	
Mean	1.14	1.06	1.02	1.00
Standard Deviation	0.04	0.04	0.04	0.04
Minimum	1.01	0.93	0.88	0.86
Maximum	1.26	1.20	1.16	1.15
			With Fixed Effects	
Mean	1.00	1.00	1.00	1.00
Standard Deviation	0.00	0.00	0.00	0.00
Minimum	1.00	1.00	1.00	1.00
Maximum	1.00	1.00	1.00	1.00

Entries are statistics (mean, std. dev., min and max) of the ratio of base year 5 to base year 1 estimated labor elasticity over 1000 repetitions. The number of firms and periods in each simulated sample are 500 and 10, respectively. The autoregressive parameters for wages and productivity are set to 0.8 in all columns.

using base year $b = 1$ over 1000 repetitions of the sample data.

Without fixed effects, the average ratio is 1.14 when there is a high serial correlation in output prices ($\rho = 0.8$) meaning that, on average, the estimated labor elasticity using base year 5 is 14 percent higher than when using base year 1. Note that, in some individual samples, the ratio goes up to 1.26. The change in estimates triggered by the change in base year is larger the higher the persistence in prices. In fact, when prices are serially uncorrelated ($\rho = 0$) the ratio is on average one – the estimates do not change on average – as expected from the discussion in Section 2.[18]

The bottom panel repeats this exercise but using fixed effects in the estimation. As expected, the choice of base year does not affect the estimated labor elasticity even when output prices are highly persistent.

In sum, this simple procedure allows us to quickly assess whether the omitted base year price can be treated as a random uncorrelated effect, as implicitly done in most empirical studies that normalize the price level in the base year. Thus, when using partially deflated revenues, a conservative

[18] But note that there are differences in individual samples: the ratio of estimates ranges between 0.86 and 1.15. This is a small sample problem. The numerical value of the estimates change when the base year is changed because the dependent variable r^*_{it} changes, even if inputs are uncorrelated with p_{ib} in the population. In this latter case, however, the estimates should be identical in expectation or when the number of time periods is large.

approach would be to always incorporate fixed effects in estimation (e.g., by using first differences as in Pozzi and Schivardi, 2016) because this guarantees that the generated estimates are invariant to the choice of base year.

4 Extensions

In this subsection we discuss two issues that are typically present when estimating production functions with panel data: the use of nominal input expenditures instead of quantities and the unbalanced nature of the typical panel dataset.

4.1 Nominal input expenditures

As with firm's output, most firm-level surveys do not report input quantities such as materials or energy (although most do report hours worked or number of employees). The surveys report, instead, the nominal expenditure incurred. It is common practice to deflate nominal expenditures by the corresponding input price to generate real quantities to be used as regressors in estimation.[19] When this input price is unobserved one can use, if available, reported input price changes or an industry-level (or any other grouping) price. This is exactly the same situation described in Section 2 (particularly in footnote 9) and, as shown next, it leads to the same type of bias described there.

To be specific, let the survey report the materials expenditure bill which, in logs, is denoted by $e_{it} = p_{Mit} + m_{it}$, and materials price changes Δp_{Mit} (p_{Mit} is log material price). The production function is given by (1) and, to focus on the missing input price, we assume that output (or the firm's specific price) is observed. We can rewrite the production function as

$$y_{it} = f(k_{it}, l_{it}, e_{it} - p_{Mit}) + \omega_{it} + \eta_{it}$$

and using the recursive formula $p_{Mit} = p_{Mib} + \sum_{s=b+1}^{t} p_{Mis}$ we have

$$y_{it} = f(k_{it}, l_{it}, e^*_{it} - p_{Mib}) + \omega_{it} + \eta_{it}$$

[19] But see Grieco et al. (2016) for an alternative approach exploiting the first order conditions of the firm's profit maximization problem.

where $e^*_{it} = e_{it} - \sum_{s=b+1}^{t} p_{Mis}$ is the "partially deflated" materials expenditure bill.

When we use e^*_{it} in estimation we are incurring a measurement error p_{Mib} which, however, is fixed over time. One could argue that in a perfectly competitive input market environment with flexible inputs − as often assumed in empirical IO − input prices are the same across firms and will be controlled for by year dummies. In this case, reported input price changes should be the same across firms which, in general, are not. It is probably more likely that variations in input prices reflects variations in input quality and in other unobservables. If input prices are persistent over time then we should expect a correlation between e^*_{it} and p_{Mib}. As in Section 2, ignoring this fixed effect may result in biased estimators of the production function parameters.

Depending on the parametric form of the production function the measurement error p_{Mib} may be linearly separable, e.g., as in a Cobb-Douglas production function. In these cases, the fixed effect will reflect *both* – output and input – prices in the base year. In other cases, the unobserved base year price level may enter non-linearly making estimation more challenging.

4.2 Unbalanced panel: differences in base year are captured by a fixed effect

Typical firm-level datasets are unbalanced panels with firms entering and exiting the sample at different points in time. A standard practice in empirical IO is to deflate all nominal variables to the same base year for all firms in the data.[20] This often requires the imputation of missing prices for periods in which firms are not in the sample. Usually this imputation is done using aggregate prices.[21] This is problematic because it implies that prices evolve equally across firms when data on prices are missing but differently when they are observed. In other words, two firms entering the sample in the same year will exhibit the same price evolution towards the base year but a different one over the years in which they appear in the sample. This approach generates the sort of bias discussed by Klette and Griliches (1996).

In this subsection we show that there is no need to assign the same base year to all firms and that we can use the first year in which a firm enters the sample as the base year for each firm, as done by

[20] For example, Eslava et al. (2004) or Foster et al. (2008) use the first year of the sample period, while Koch et al. (2021) use the middle year.

[21] For example, the missing price changes before a firm enters the sample - or after it exits - are imputed by using the average price changes in its industry (and sometimes geographic location) in each of the missing years.

Jaumandreu and Lin (2018). In this case, differences in base years prices across firms are captured by differences in the firm fixed effects.[22]

Consider two firms entering the sample at different dates, firm i at $t = b$ and firm j at $t = b' > b$, with revenues $R_{it} = P_{it}Y_{it}$ and $R_{jt} = P_{jt}Y_{jt}$. If the base year is the first year the firm enters the sample deflated revenues for each firm are given by

$$R^*_{it} = \frac{P_{it}Y_{it}}{P_{it}/P_{ib}} = P_{ib}Y_{it} \quad \text{and} \quad R^*_{jt} = \frac{P_{jt}Y_{jt}}{P_{jt}/P_{jb'}} = P_{jb'}Y_{jt}$$

and, in logs,

$$r^*_{it} = p_{ib} + y_{it} \quad \text{and} \quad r^*_{jt} = p_{jb'} + y_{jt}$$

and following (4) we have that

$$r^*_{it} = f(k_{it}, l_{it}, m_{it}) + p_{ib} + \omega_{it} + \eta_{it} \quad \text{and} \quad r^*_{jt} = f(k_{jt}, l_{jt}, m_{jt}) + p_{jb'} + \omega_{jt} + \eta_{jt}$$

which clearly shows that differences in base year prices are captured by differences in the fixed effect. That is, if firms i and j are identical at time t then $r^*_{it} - r^*_{jt} = \left(r_{it} - \sum_{s=b+1}^{t}(\Delta p_{is})\right) - \left(r_{jt} - \sum_{s=b'+1}^{t}(\Delta p_{js})\right) = -\sum_{s=b+1}^{b'} \Delta p_{is} = p_{ib} - p_{jb'}$ reflects the accumulated price changes between the two base years b and b', which is nothing more than the rate of price change between the two entry dates. This difference, however, is constant over time and is captured by the difference in firm fixed effects.

In sum, in unbalanced panels we do not want to use a common base year across firms because the price changes imputations are likely to introduce an omitted variable bias. Incorporating fixed effects in the revenue production function allows us to deflate each firm's revenues to their respective entry date. Essentially, the firm fixed effect is picking up the unobserved price level in *any* base year.

5 Empirical application I

In this Section we used real firm level data to illustrate how the choice of base year affects the estimated production function elasticities and how this problem is avoided by the introduction of firm-

[22] We illustrate this point using the first year in which the firm appears in the sample but the same argument applies to any year in which the firm is present in the sample.

specific fixed effects into the estimation.

This empirical analysis uses data from the Spanish firm-level database Encuesta Sobre Estrategias Empresariales (ESEE), which has been widely used to estimate production functions and related issues. In this illustrative application we use the subsample of small and medium sized firms (less than 200 workers) that entered the sample between 1991 and 1996 and stayed in the panel for at least four consecutive years. We end the sample period in 1999.[23]

The ESEE survey does not report price levels but it does report the rate of change of its prices. More precisely, firms communicate the yearly average rate of change in the prices in at most five separate markets. Based on this information, a firm-specific output price index is constructed as a Paasche-type index where the base year is normalized to unity. A similar price index is constructed for materials using reported price changes of the main material inputs.

The theoretical part in Section 2 considers a firm producing and selling a single product but the data include firms producing different products which are sold in different markets. This issue is relevant to all production function estimation methods. Although there are methodological advances in estimating production functions with multiple products, these approaches require disaggregated data not readily available (see Pakes 2021 for a brief summary). In our case, we would need, in addition, an understanding of the type of price changes reported, i.e., are price changes reported for specific product-maket combinations? In our sample of small firms, 82 percent of them produce a single 3-digit NACE product, while 12 percent of them produce in two different 3-digit NACE products. In practice, we therefore treat these firms as producing a single product and the price change used is the a weighted average of the reported price changes in at most five markets.

A message of the simulation results in Table 2 is that the impact of ignoring the unobserved base year price level in estimation depends on the persistence of the output price process. The higher the serial correlation in prices, the stronger the omitted variable bias. In our sample data, the simple autocorrelation coefficient of the constructed output price index is between 0.92 and 0.97, depending on the sector, and declines, as expected, to 0.6-0.73 when controlling for year and firm effects via a regression.[24] These results suggest that output prices are highly persistence raising the concern that

[23] See Appendix C for a description of the dataset used in this application.

[24] Similarly, Foster et al. (2008) report that output prices are highly persistent in their sample of US manufacturing plants

the normalization used in constructing the price indices may generate a significant omitted variable bias.

We use OLS to estimate a Cobb-Douglas production function with and without firm fixed effects. This estimator is not consistent because of omitted productivity but, as argued at the end of Section 2 and shown in the simulations in Table 3, the probability limit of the OLS estimator should be invariant to the choice of base year if the price in the base year is uncorrelated with input choices. Because OLS is a simple and easy estimator to compute, it is natural to use it to evaluate the adequacy of normalizing the price in the base year across firms.

We estimate the production function twice using the same sample data but changing the base year of both output and material price indices. The initial base year, denoted by b, is the calendar year *prior to the first year* in which the firm is observed in the sample. Thus, each firm has a different base year because they enter the sample in different years.[25] We then change the base year to $b' = b + 3$. Thus, a firm first observed in year $1991 \leq \tau \leq 1996$ reports its price change between $\tau - 1$ and τ and we set $b = \tau - 1$. Because firms are in the sample at least 4 years we change the base year to $b' = \tau + 2$ and use the reported price change in $\tau + 3$. For example, firms first observed in 1993 initially have base year 1992 which is then changed to 1995.

In Table 3 we present the ratio of the estimated elasticities using base year $b + 3$ to the estimates using base year b. This ratio should be equal to one if the estimates are not affected by the change in base year.

Using OLS *without fixed effects* gives ratios that in most cases deviate from one, particularly for labor and capital. This means that the estimates change when the base year changes. For example, the coefficients' ratio for labor in the Food and Beverages sector is 1.43, which means that the estimated labor elasticity increase by 43 percent when the base year changes from b to $b + 3$. For capital, the coefficients' ratio is 0.85 meaning that the estimated capital elasticity decreased by 15 percent. Many of these changes are quantitatively important. It is not a question of whether these changes are large enough to reject a null hypothesis of no correlation between the base year prices and the

with implied annual autocorrelation values of roughly 0.75 to 0.80.

[25] The survey reports at year t the rate of output price change between $t - 1$ and t. Most empirical papers use a common base year which requires imputation of missing prices for periods in which firms are not in the sample (see subsection 4.2).

Table 3: Impact of change in base year on estimated coefficients

	Ratio of estimated coefficients between base years $b+3$ and b					
	OLS			OLS with fixed effects		
	Materials	Labor	Capital	Materials	Labor	Capital
Textiles	1.00	1.12	0.92	1	1	1
Food and Beverages	0.99	1.43	0.85	1	1	1
Chemicals	0.98	1.06	0.94	1	1	1

Entries are ratios of the estimated parameters based on base year $b+3$ to the estimates based on year b, where b is the year prior to entry for each firm. The base years of both output and materials price indices are changed.

input regressors. The fact that the estimates change *at all* when the base year is changed should be surprising. Choosing the base year for the price index is not an innocuous decision.

These findings, which are in line with the simulation results in subsection 3.2.2, tell us that treating the unobserved base year price as an uncorrelated component of the error term or, equivalently, normalizing it to the same value across all firms, is not appropriate for these data. Ignoring the heterogeneity in base year prices is troublesome because the estimated parameters are sensitive to the choice of base year.

This undesirable situation can be avoided by incorporating firm fixed effects in the estimation of production functions when revenues are partially deflated. As shown in Table 3, the coefficient ratios when using OLS *with fixed effects* are exactly one for all inputs and in all sectors. However, incorporating fixed effects while also accounting for unobserved productivity requires a more sophisticated estimator. One such estimator is presented in the next Section.

6 The GNR production function estimator

In this Section we briefly sketch the production function estimator developed by Gandhi et al. (2020; henceforth GNR). This estimator is based on the proxy-variable/control approach initiated by Olley and Pakes (1996) and Levinsohn-Petrin (2003). Originally, the proxy-variable approach was developed to estimate parametric production functions that did not typically allow for firm-level fixed effects (Ackerberg et al., 2015; Ackerberg, 2021). GNR offer a nonparametric procedure that identifies the input elasticities of an unknown output production function and allows for firm specific fixed-effects which, as shown in Section 2, are necessary to account for unobserved prices in the base year.

We first focus on the case where only output is not observed and we show how the unobserved base year output price level is an additive fixed effect in the dependent variable that can be easily incorporated into the original GNR approach.

The production function is given by equation (1), where $f(k_{it}, l_{it}, m_{it})$ is an unknown function, and ω_{it} follows a Markov process, $\omega_{it} = h(\omega_{it-1}) + \epsilon_{it}$ with unknown function $h(\cdot)$ and ϵ_{it} is I.I.D. Materials is a flexible input chosen at time t after observing ω_{it} but before the production shock η_{it} is realized. Capital and labor – the so-called dynamic inputs – are both predetermined at time t, i.e., they are chosen given information at time $t-1$, ω_{it-1} and η_{it-1}.

As in most proxy variable methods, the GNR approach is a two stage procedure. In the first stage the elasticity of the flexible input is identified from the link between the first order condition for materials and materials' expenditure share. The *share regression* used to identify the elasticity of the flexible input m_{it} is derived from the log of the firm's first order conditions,

$$s_{it} = \underbrace{ln(Ee^{\eta_{it}}) + ln\frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it})}_{\ln D^{\mathcal{E}}(k_{it}, l_{it}, m_{it})} - \eta_{it} \tag{9}$$

where $s_{it} = ln\frac{P_{Mit}M_{it}}{P_{it}Y_{it}}$ is the log of the materials expenditure share, and P_{Mit} is the price of materials.

When the production function is unknown, GNR approximate the log of the elasticity by a polynomial expansion. Thus, equation (9) may be estimated by non-linear least squares on (k_{it}, l_{it}, m_{it}) and we can then identify the shock $\eta_{it} = s_{it} - \ln D^{\mathcal{E}}(k_{it}, l_{it}, m_{it})$ and the constant $ln(Ee^{\eta_{it}})$ from which we obtain the flexible input (log) elasticity of production $ln\frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it})$.

In the second stage, the dynamic inputs' elasticities are identified by solving a partial differential equation defined by the flexible input elasticity. The idea is that we integrate back the derivative of the production function in the flexible input to get the production function up to a constant of integration,

$$\int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} = f(k_{it}, l_{it}, m_{it}) + \mathcal{C}(k_{it}, l_{it})$$

where $\mathcal{C}(k_{it}, l_{it})$ is an unknown constant of integration that depends only on the dynamic inputs.

Substituting for the production function and rearranging terms we have that productivity is

$$\omega_{it} = y_{it} - \int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} - \eta_{it} + \mathcal{C}(k_{it}, l_{it}) \tag{10}$$

Let us assume, for the moment, that output y_{it} is observable. Define

$$\mathcal{R}_{it} \equiv y_{it} - \int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} - \eta_{it} \tag{11}$$

which is an observable quantity since the last two terms are obtained in the first stage.

We therefore have $\omega_{it} = \mathcal{R}_{it} + \mathcal{C}(k_{it}, l_{it})$ where only the last term is unknown. Rearranging and substituting for ω_{it} with $\omega_{it} = h(\omega_{it-1}) + \epsilon_{it}$ results in

$$\mathcal{R}_{it} = -\mathcal{C}(k_{it}, l_{it}) + h\left(\mathcal{R}_{it-1} + \mathcal{C}(k_{it-1}, l_{it-1})\right) + \epsilon_{it}$$

which can be used to non-parametrically estimate the production function because all the observable variables are predetermined with respect to the productivity innovation ϵ_{it}.

GNR approximate the unknown functions $h(\cdot)$ and $\mathcal{C}(\cdot)$ by means of polynomial expansions. We call this procedure the *nonparametric GNR* estimator.

When the firm's output y_{it} is partially observed because the base year price level is unobserved, $y_{it} = r_{it} - p_{it} = r^*_{it} - p_{ib}$, equation (10) becomes

$$\begin{aligned}
\omega_{it} &= r^*_{it} - p_{ib} - \int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} - \eta_{it} + \mathcal{C}(k_{it}, l_{it}) \\
&= \mathcal{R}^*_{it} - p_{ib} + \mathcal{C}(k_{it}, l_{it})
\end{aligned}$$

where $\mathcal{R}^*_{it} = r^*_{it} - \int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} - \eta_{it}$ is observed after the first stage.

Following the same steps as before, implies

$$\mathcal{R}^*_{it} = p_{ib} - \mathcal{C}(k_{it}, l_{it}) + h\left(\mathcal{R}^*_{it-1} - p_{ib} + \mathcal{C}(k_{it-1}, l_{it-1})\right) + \epsilon_{it}$$

so that the firm effect p_{ib} enters non-parametrically through the unknown function $h(\cdot)$.

GNR address this problem by making an additional assumption on the productivity process, namely, that $h(\cdot)$ is linear so that productivity follows a linear AR(1) process:

$$\omega_{it} = h(\omega_{it-1}) + \epsilon_{it} = \rho_\omega \omega_{it-1} + \epsilon_{it}.$$

This implies that p_{ib} enters in a linearly separable fashion,

$$\mathcal{R}^*_{it} = p_{ib}\,(1-\rho_\omega) - \mathcal{C}(k_{it}, l_{it}) + \rho_\omega \left(\mathcal{R}^*_{it-1} + \mathcal{C}(k_{it-1}, l_{it-1})\right) + \epsilon_{it} \tag{12}$$

and $p_{ib}\,(1-\rho_\omega)$ is now the fixed effect.

GNR approximate the unknown function $\mathcal{C}(\cdot)$ by means of a polynomial expansion. We call this procedure the *nonparametric fixed effect GNR* (*nonparametric FE GNR*) estimator.

In sum, allowing for fixed effects in GNR comes at a price: we need to assume a linear AR(1) process for the unobserved productivity whereas no such assumption is necessary when ignoring the fixed effects. The reason for this additional assumption is the need to avoid non-parametric functions of the unobserved fixed effect.

An important remark is that the lack of information on *output* price levels only affects the second stage of the GNR procedure. The flexible input elasticity estimator in the first stage is not functionally dependent on the output price level and, therefore, is not affected by changes in the base year of the output price.[26]

Finally, we adapt the GNR approach to the case of a Cobb-Douglas production function which is the most common specification used in the estimation of production functions. When $f(k_{it}, l_{it}, m_{it})$ is Cobb-Douglas the first stage simplifies to $s_{it} = ln(Ee^{\eta_{it}}) + \log\beta_M - \eta_{it}$. Let $\mu_M = ln(Ee^{\eta_{it}}) + \log\beta_m$ be the population mean of s_{it}. Then $\eta_{it} = \mu_M - s_{it}$ and

$$\log\beta_M = \mu_M - \log(Ee^{\mu_M - s_{it}}) \Longrightarrow \beta_M = e^{\mu_M} - Ee^{\mu_M - s_{it}} \tag{13}$$

which can be consistently estimated using the sample mean of s_{it}. Notice that the (non)observability of the output price does affect the estimation of β_M.

[26] However, as we will see later on, it may still be affected by the choice of base year for the input price when *materials* are partially deflated.

In the second step, the estimated equation (12) becomes[27]

$$\mathcal{R}^*_{it} = p_{ib}\left(1-\rho_\omega\right) + \beta_K k_{it} + \beta_L l_{it} + \rho_\omega\left(\mathcal{R}^*_{it-1} - \beta_K k_{it-1} - \beta_L l_{it-1}\right) + \epsilon_{it} \tag{14}$$

which is a dynamic panel data model with fixed effects subject to parameter restrictions. We refer to this specification as the *GNR (FE) Cobb-Douglas* estimator.

The Cobb-Douglas specification in conjunction with the linear AR(1) assumption for ω_{it} have another important advantage: they can easily accommodate the use of partially deflated materials, whereas more general production functions do not. To see this, recall from subsection 4.1 that when m_{it} is not observed and we use reported materials price changes to deflate nominal materials expenditures we are using an error-ridden measure of the quantity of materials $e^*_{it} = m_{it} + p_{Mib}$. In the general case, the use of e^*_{it} instead of m_{it} introduces a fixed effect in the non-parametric production function appearing in (11). Under Cobb-Douglas, however, $\mathcal{R}^*_{it} = r^*_{it} - \beta_M m_{it} - \eta_{it} = r^*_{it} - \beta_M e^*_{it} - \eta_{it} + \beta_M p_{Mib}$. Furthermore, the linear AR(1) assumption implies that $\mathcal{R}^*_{it-1}$ enters linearly so that the fixed effect in (14) when using partially deflated materials expenditures becomes $\left(p_{ib} - \beta_M p_{Mib}\right)\left(1-\rho_\omega\right)$.

In the next section we apply the GNR approach to the Spanish data set.

7 Empirical application II

As discussed above, the GNR approach is a proxy method that allows for fixed effects in the production function. Because the unobserved base year price level can be treated as a firm-specific fixed effect, we use the GNR estimator, with and without fixed effects, to examine the sensitivity of the estimated input elasticities to changes in the base year.

This is the same type of exercise done in Section 5 using the biased OLS estimator. We estimate the production function using two different base years, b and $b' = b + 3$, and compare the estimated elasticities. As in Section 5, base year b is the calendar year prior to the the first year the firm is observed in the sample and, because we require firms to be in the sample for at least four consecutive

[27] After substituting $f(k_{it}, l_{it}, m_{it}) = \beta_K k_{it} + \beta_L l_{it} + \beta_M m_{it}$ and noting that $\int \frac{\partial}{\partial m_{it}} f(k_{it}, l_{it}, m_{it}) dm_{it} = \beta_M m_{it}$ implies $\mathcal{C}(k_{it}, l_{it}) = -\left(\beta_K k_{it} + \beta_L l_{it}\right)$.

years, we set $b' = b + 3$.

In Table 4 we examine three cases. First, in the baseline case (i), output and materials expenditures are both deflated using base year b and we present the estimated parameters for each of the three sectors analyzed. In the second case (ii), we deflate output using base year $b' = b + 3$, while materials expenditures remain deflated using base year b, i.e., we change the base year only for the output price index. Finally, in case (iii), we change the base year for both output and materials price indeces by using base year $b' = b + 3$ for both. In these last two cases we show the ratio of the estimates in the two cases to the estimates in the baseline case. This ratio indicates the sensitivity of the estimates to the choice of base year.[28]

As explained in Section 6, the estimated material elasticity is not affected by the change in base year and, indeed, the ratio between the different base-year estimates is always one. This is not the case for labor an capital when fixed effects are ignored (panels A and C). The changes generated by choosing different base years can be quantitatively important. For example, when using the nonparametric GNR (panel A) and changing the base year of the output price index only, the mean labor elasticity increases by 60 percent in Textiles but does not change in Chemicals. There is no clear pattern to the changes: sometimes the elasticities increase while at other times they decrease. In general, and as expected, when the base years of both price indices change, the change in estimated elasticities is larger, sometimes considerably so (e.g., labor elasticity in Textiles now increases by 153 percent!).

When fixed effects are included (panels B and D) the ratio is, as expected, essentially one when only the base year for the output price is changed: the choice of base year does not affect the estimated elasticities.[29] However, when the base years of both price indices change, the nonparametric FE GNR estimator does change by about 6-13 percent. The reason for these changes is that the unobserved base year materials price enters the estimated equation non-linearly and therefore cannot be fully accounted for by linear fixed effects. An estimator that handles non-linear fixed effects should be invariant to this problem. The results also show that, in general, the sensitivity to the choice of base year is an order of magnitude smaller when using fixed effects than when ignoring them (compare

[28] Here we do not explicitly compare among the different estimators. We do this in the next subsection where we estimate aggregate productivity using the different estimators.

[29] The ratio is exactly one for the Cobb-Douglas FE GNR estimator, and very close to one for the nonparametric FE GNR. We attribute this small discrepancy from unity to the more complex numerical procedures used by GNR in the nonparametric case.

Table 4: The impact of changing the base year on production function estimates

	Textiles			Food and Beverages			Chemicals		
	Elasticity	Ratio		Elasticity	Ratio		Elasticity	Ratio	
Choice of the base year for:	i	ii	iii	i	ii	iii	i	ii	iii
Output price index	b	b'	b'	b	b'	b'	b	b'	b'
Materials price index	b	b	b'	b	b	b'	b	b	b'
				Panel A: Nonparametric GNR					
Materials	0.55	1.00	1.00	0.66	1.00	1.00	0.65	1.00	1.00
Labor	0.05	1.60	2.53	0.09	1.17	1.91	0.31	1.00	0.99
Capital	0.14	0.99	1.12	0.17	0.86	0.72	0.04	1.24	1.76
				Panel B: Nonparametric FE GNR					
Labor	0.63	1.00	1.06	0.29	1.00	0.94	0.05	0.99	1.13
Capital	0.23	1.00	1.05	0.32	1.00	1.01	0.37	1.00	0.99
				Panel C: Cobb - Douglas NR					
Materials	0.32	1.00	1.00	0.60	1.00	1.00	0.62	1.00	1.00
Labor	0.65	0.96	0.95	0.24	1.02	1.02	0.33	0.99	0.97
Capital	0.20	1.09	1.12	0.15	1.09	1.16	0.10	0.98	1.10
				Panel D: Cobb - Douglas FE GNR					
Labor	0.77	1.00	1.00	0.06	1.00	1.00	0.03	1.00	1.00
Capital	0.29	1.00	1.00	0.37	1.00	1.00	0.45	1.00	1.00

Case i, the baseline case, is when the base year b is the calendar year prior to the first year in which the firm is observed in the sample. The entries are mean elasticities over all firms and years in the sector.
Case ii is when the base year for the output is $b' = b + 3$ but materials was left at b.
Case iii output and materials base year are b'. In case ii and iii we present the ratio of estimates wrt i.
In panel C (GNR Cobb-Douglas) in all sectors except Textiles, the polynomial expansion of the unknown function was of order 2 and contemporary labor was used in the moment condition for labor elasticity. In Textiles, the polynomial expansion was of order three and labor lagged one period was used.

panels A and B).

The big advantage of the GNR Cobb-Douglas estimator now comes to the fore. As explained at the end of section 6, a Cobb-Douglas assumption (jointly with a linear AR(1) productivity process) can easily account for the unobserved base year material price. The results confirm this by showing that the ratio of the estimated coefficients is always exactly one. Thus, if the Cobb-Douglas and linear AR(1) productivity process assumptions are justified, there is a simple way to use the data on price changes: construct firm-specific price indices but also incorporate fixed effects in the estimation. If these assumptions are not adequate, and one uses FE GNR, we then suggest reporting how sensitive the results are to the choice of base year.

In sum, the normalization of the unobserved base year price level is not an innocuous decision. The empirical results confirm the premise of this paper that the unobserved base year output price introduces an omitted variable problem. The production function elasticity estimates are sensitive to the choice of base year when the unobserved base year price levels are omitted, i.e., when normalized to zero. Furthermore, this omitted variable problem can be overcome to a large extent by incorporating firm fixed effects in the estimation of the production function.

7.1 Productivity

In this subsection we first show how estimated productivity depends on the choice of base year, and then how it changes across the various estimators. We already know that the estimated production function is not invariant to the choice of base year when fixed effects are ignored. Here we examine the quantitative implications of this on aggregate productivity.

Using (4) and the estimated production function we compute residuals $\widehat{u}_{it}$ to estimate firm-level productivity levels, namely,

$$\begin{aligned}
\widehat{u}_{it} &\equiv r^*_{it} - \widehat{f(k_{it}, l_{it}, m_{it})} - \widehat{\eta}_{it} \\
&= y_{it} + p_{it} - \sum_{s=b+1}^{t} (\Delta p_{is}) - \widehat{f(k_{it}, l_{it}, m_{it})} - \widehat{\eta}_{it} \\
&= y_{it} - \widehat{f(k_{it}, l_{it}, m_{it})} - \widehat{\eta}_{it} + p_{ib} \\
&\equiv \hat{\omega}_{it} + p_{ib}
\end{aligned}$$

where an " $\widehat{}$ " means an estimator of the underlying variable and $\widehat{\eta}_{it}$ is obtained in the first step of the GNR procedure .

Clearly, this residual includes the unobserved output price in the base year and is therefore not a clean measure of productivity ω_{it}. We compute $\widehat{u}_{it}$ using the four estimators described in Section 7, aggregate up to the sectorial level using revenue shares, and compute growth rates,

$$\begin{aligned}\hat{\Phi}_t - \hat{\Phi}_{t-1} &= \sum_i s_{it}\hat{u}_{it} - \sum_i s_{it-1}\hat{u}_{it-1} \\ &= \sum_i s_{it}\left(\hat{u}_{it} - \hat{u}_{it-1}\right) + \sum_i \hat{u}_{it-1}\left(s_{it} - s_{it-1}\right)\end{aligned}$$

where $s_{it} = \frac{R_{it}}{\sum_{i=1}^{N_t} R_{it}}$ is the revenue (R_{it}) share of firm i at time t.

This measure of productivity growth, however, is not invariant to the choice of base year for two reasons. First, the within-firm growth component $\hat{u}_{it} - \hat{u}_{it-1} = \hat{\omega}_{it} - \hat{\omega}_{it-1}$ is determined by the estimated production function which, as previously shown, depends on the choice of base year when fixed effects are not accounted for in the estimation. This problem can then be easily avoided by using estimators that incorporate such fixed effects. Second, the term $\hat{u}_{it-1}\left(s_{it} - s_{it-1}\right) = \left(\hat{\omega}_{it-1} + p_{ib}\right)\left(s_{it} - s_{it-1}\right)$ depends directly on p_{ib} and will therefore over (under) estimate the reallocation component to the extent that changes in market shares are positively (negatively) correlated with the base year price level. Because p_{ib} is unobserved this bias cannot be avoided.[30]

We remark that even if p_{ib} can be treated as a random component – uncorrelated with inputs at time t – and fixed effects are therefore not required for consistent estimation, the reallocation component of productivity growth is still biased as long as prices in the base year are correlated with market shares changes, i.e., when $\sum_i p_{ib}\left(s_{it} - s_{it-1}\right) \neq 0$.

Table 5 shows $\hat{\Phi}_t - \hat{\Phi}_{t-1}$ averaged over the eight years 1992-1999. Even when using the nonparametric GNR with fixed effects the estimated productivity varies considerably with the choice of base year. For example, in the Food and Beverages sector, the average annual productivity growth is -0.4, -1.06 and 0.36 percent depending on the base year chosen. The Chemical sector, however, is less sensitive to the choice of the base year. Again, choosing the base year is not an innocuous decision because it has a measurable impact on estimated productivity growth. Table 5 also allows us to compare across estimators. Focusing, for example, on the baseline normalization (case i) and on the Food and Beverages secotr we see that $\hat{\Phi}_t - \hat{\Phi}_{t-1}$ varies considerably across the four estimators in both, magnitude and direction. For example, the average growth rate during the period is 0.92 percent per year when using the nonparametric GNR estimator, but -0.4 percent per year when using

[30] Unless, of course, p_{ib} is uncorrelated with market share changes.

Table 5: Annual aggregate productivity growth 1991-1999 (%)

	Textiles			Food and Beverages			Chemicals		
Choice of the base year for:	i	ii	iii	i	ii	iii	i	ii	iii
Output price index	b	b'	b'	b	b'	b'	b	b'	b'
Materials price index	b	b	b'	b	b	b'	b	b	b'
Nonparametric GNR	-0.29	-0.30	0.05	0.92	0.53	2.13	3.76	3.43	3.11
Nonparametric FE GNR	-1.22	-1.21	-0.87	-0.40	-1.06	0.36	-2.52	-2.59	-2.37
GNR Cobb - Douglas	0.52	0.39	0.46	1.73	0.93	1.65	4.09	4.04	4.05
GNR FE Cobb - Douglas	-0.30	-0.20	-0.13	-0.20	-0.80	0.02	0.26	0.19	0.32

Entries are the simple average of the annual aggregate productivity growth for the eight years between 1992 and 1999. The annual aggregate productivity growth is the difference between period t and period t-1 weighted average (log) productivity level. Case i, the baseline case, is when the base year b is the calendar year prior to the first year in which the firm is observed in the sample. In case ii(iii) we change the base year for the output (output and materials) price index to $b' = b + 3$.

the FE GNR estimator giving a 1.32 pp difference between them. A similar result holds for the Cobb-Douglas GNR estimator. Thus, accounting for fixed effects matters for the ultimate goal of estimating aggregate productivity growth.

Table 5 also allows us to compare across estimators. Focusing, for example, on the baseline normalization (case i) and on the Food and Beverages secotr we see that $\hat{\Phi}_t - \hat{\Phi}_{t-1}$ varies considerably across the four estimators in both, magnitude and direction. For example, the average growth rate during the period is 0.92 percent per year when using the nonparametric GNR estimator, but -0.4 percent per year when using the FE GNR estimator giving a 1.32 pp difference between them. A similar result holds for the Cobb-Douglas GNR estimator. Thus, accounting for fixed effects matters for the ultimate goal of estimating aggregate productivity growth.

The literature documents large and persistent productivity differences across firms, even within narrowly defined industries (Syverson, 2011). The purpose of many theoretical and empirical studies is to understand the reasons for this observed dispersion. Here we want to show that accounting for fixed effects also matters for measuring productivity dispersion.

We compute the standard deviation of $\widehat{u}_{it}$ across firms in the same industry and year. Clearly, this measure of dispersion reflects not only the dispersion of true productivity ω_{it} but also that of

Table 6: Productivity dispersion

Sector	**Estimator**	1991	1992	1993	1994	1995	1996	1997	1998	1999	Total
Textiles	Nonparametric GNR	0.26	0.26	0.26	0.25	0.26	0.26	0.28	0.28	0.28	**0.27**
	Nonparametric FE GNR	0.45	0.46	0.46	0.48	0.50	0.50	0.52	0.53	0.52	**0.49**
	GNR Cobb-Douglas	0.87	0.85	0.87	0.90	0.86	0.94	0.94	0.96	0.90	**0.90**
	GNR FE Cobb-Douglas	0.90	0.87	0.88	0.90	0.86	0.94	0.93	0.96	0.89	**0.90**
Food & Beverages	Nonparametric GNR	0.16	0.14	0.17	0.21	0.23	0.24	0.25	0.25	0.26	**0.21**
	Nonparametric FE GNR	0.37	0.35	0.35	0.41	0.43	0.44	0.46	0.46	0.45	**0.41**
	GNR Cobb-Douglas	0.36	0.38	0.37	0.41	0.43	0.44	0.43	0.43	0.44	**0.41**
	GNR FE Cobb-Douglas	0.42	0.41	0.40	0.43	0.44	0.45	0.46	0.45	0.46	**0.44**
Chemicals	Nonparametric GNR	0.08	0.10	0.14	0.16	0.18	0.19	0.20	0.22	0.23	**0.17**
	Nonparametric FE GNR	0.33	0.30	0.31	0.34	0.36	0.37	0.38	0.40	0.40	**0.35**
	GNR Cobb-Douglas	0.26	0.28	0.28	0.30	0.29	0.30	0.31	0.34	0.31	**0.30**
	GNR FE Cobb-Douglas	0.47	0.50	0.46	0.49	0.43	0.46	0.45	0.44	0.39	**0.45**

Entries are standard deviations of productivity across firms in the same industry and year.
Base year used is b, the year prior to entry into the sample (case i).

output prices in the base year p_{ib}. Our focus, however, is on the comparison of this dispersion across the four estimators presented above. Table 6 presents the cross-sectional dispersion for each industry and year. Comparing the nonparametric GNR estimator with and without fixed effects it is apparent that the estimated dispersion of productivity (and p_{ib}) is much larger when fixed effects are incorporated in the estimation. Thus, accounting for fixed effects also matters for measuring productivity dispersion.

8 Summary and conclusions

A problem in production function estimation using firm level data is that physical quantities are typically unavailable. Instead, most studies use revenues and expenditures on inputs deflated by industry-level prices. As shown by Klette and Griliches (1996), doing this generates an omitted variable bias.

In this paper we show that using firm-specific price indices constructed from observed price changes to deflate nominal revenue and/or input expenditures also generates an omitted variable problem. This bias is a direct result of the arbitrary normalization of prices in the base year used to construct the price indices.

This omitted variable, however, is constant over time and varies across firms so that it can be accounted for by incorporating fixed effects into the estimation of production functions. We apply recent production function estimators that allow for FE and examine the implications of doing so in a popular Spanish firm-level data set. We find that using FE matters both for the estimation of elasticities and for the estimation of aggregate productivity growth and productivity cross-sectional dispersion.

A message from this paper is than when constructing firm-level price indices based on price changes only, the choice of base year matters. If FE cannot be incorporated into the estimation, then researchers should report the sensitivity of the results to the choice of base year.

References

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

[2] Ackerberg, Daniel, 2021 "Comment on Olley and Pakes-style Production Function Estimators with Firm Fixed Effects" Oxford Bulletin of Economics and Statistics, 83 (3): 836-840. doi:10.1111/obes.12406.

[3] Bartelsman, Eric, J., and Mark Doms, 2000 "Understanding Productivity: Lessons from Longitudinal Microdata" Journal of Economic Literature, 38(3): 569–594.

[4] Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, 2021 "Some Unpleasent Markup Arithmetic. Production Function Elasticities and their Estimation from Production Function" Journal of Monetary Economics, 121(1): 1–14. https://doi.org/10.1016/j.jmoneco.2021.05.004.

[5] Carlsson, Mikael, Julián Messina, and Oskar Nordström Skans, 2021 "Firm-level Shocks and Labour Flows" The Economic Journal, 131(634): 598–623.

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

[7] De Loecker, Jan, 2021 "Comment on (Un)pleasant ... by Bond et al. (2020)" Journal of Monetary Economics, 121: 15–18.

[8] Decker, Ryan A., John Haltiwanger, Ron S. Jarmin, and Javier Miranda 2020 "Changing Business Dynamism and Productivity: Shocks versus Responsiveness" American Economic Review, 110(12): 3952–90.

[9] Dolado, J., Salvador Ortigueira, and Rodolfo Stucchi, 2016 "Does dual employment protection affect TFP? Evidence from Spanish manufacturing firms" *SERIEs*, 7(4): 421–459.

[10] Eslava, Marcela, John Haltiwanger, Adriana Kugler, and Maurice Kugler, 2004 "The Effects of Structural Reforms on Productivity and Profitability Enhancing Reallocation: Evidence from Colombia" Journal of Development Economics, 75(2): 333–371.

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

[12] Gandhi, Amit and Salvador Navarro, and David A. Rivers, 2020 "On the Identification of Gross Output Production Functions" Journal of Political Economy, 128(8): 2973–3016.

[13] Grieco, Paul L.E., Shengyu Li and, Hongsong Zhang, 2016 "Production Function Estimation with Unobserved Input Price Dispersion" International Economic Review, 57: 665–690.

[14] Jaumandreu, Jordi, and Shuheng Lin, 2018 "Prices Under Innovation: Evidence from Manufacturing Firms" CEPR Discussion Paper No. DP13146, Available at SSRN: https://ssrn.com/abstract=3244520.

[15] Klette, Tor Jakob, and Zvi Griliches, 1996 "The Inconsistency of Common Scale Estimators when Output Prices are Unobserved and Endogenous" Journal of Applied Econometrics, 11(4): 343–361.

[16] Koch, Michael, Ilya Manuylov, and Marcel Smolka, 2021 "Robots and Firms" The Economic Journal, 131(638): 2553–2584.

[17] Levinsohn, J., and Anil Petrin, 2003 "Estimating Production Functions Using Inputs to Control for Unobservables" The Review of Economic Studies, 70(2): 317–341.

[18] Mairesse, Jacques, and Jordi Jaumandreu, 2005 "Panel-data Estimates of the Production Function and the Revenue Function" Scandinavian Journal of Economics, 107(4): 651–672.

[19] Olley, G. Steven, and Ariel Pakes 1996 "The Dynamics of Productivity in the Telecommunications Equipment Industry" Econometrica, 64(6): 1263–97.

[20] Pakes, Ariel, 2021 "A Helicopter Tour of Some Underlying Issues in Empirical Industrial Organization" Annual Review of Economics, 13: 397–421.

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

[22] Roberts, Mark, and Dylan Supina, 2000 "Output Price and Markup Dispersion in Producer Micro Data: The Roles of Producer Heterogeneity and Noise" in Advances in Applied Microeconomics, Vol 9, Industrial Organization, Michael Baye (ed.), JAI Press.

[23] Smeets, Valerie, and Frederic Warzynski, 2013 "Estimating Productivity with Multi-product firms, Pricing Heterogeneity and the Role of International Trade" Journal of International Economics, 90(2): 237–244.

[24] Syverson, Chad, 2011 "What Determines Productivity?" Journal of Economic Literature, 49 (2): 326–65.

A A monopolistic competition model

We present an alternative model of firm behavior to illustrate how prices in the base year can be correlated with input choices in future years when prices are endogenous. For simplicity, we assume a Cobb-Douglas production function with one variable input, labor (L). The production function of firm i is given by (5),

$$Y_{it} = L_{it}^{\alpha} e^{\omega_{it} + \eta_{it}} \tag{15}$$

with $\alpha \leq 1$.

The firm observes ω_{it} and η_{it} when making the labor input decision. This timing assumption is made to avoid discrepancies between quantity demanded and quantity produced and therefore differs from the timing assumption in Section 3.1. The cost function is

$$C_{it} = C\left(Y_{it}, \omega_{it}, \eta_{it}\right) = P_{Lit} \left(\frac{Y_{it}}{e^{\omega_{it} + \eta_{it}}} \right)^{\frac{1}{\alpha}}$$

where P_{Lit} is the wage in year t. Marginal cost is decreasing in output

$$\frac{\partial C\left(Y_{it}, \omega_{it}, \eta_{it}\right)}{\partial Y_{it}} = \frac{C_{it}}{\alpha Y_{it}}$$

The competitive environment is one in which firms sell a differentiated product in monopolistic competition. We assume the following (residual) demand function for the product of firm i

$$P_{it} = \mu_{it} Y_{it}^{\phi - 1} \tag{16}$$

as in Decker et al. (2020), where P_{it} is the price of product i in period t and μ_{it} is an exogenous demand shifter. We assume $\phi \leq 1$.

At the beginning of period t, ω_{it}, η_{it} and μ_{it} are observed by the firm when the firms chooses its price. Assuming monopolistic competition, the price that maximizes per-period profits is given by equating marginal cost to marginal revenue:

$$\mu_{it} \phi Y_{it}^{\phi - 1} = \frac{P_{Lit}}{\alpha Y_{it}} \left(\frac{Y_{it}}{e^{\omega_{it} + \eta_{it}}} \right)^{\frac{1}{\alpha}} \tag{17}$$

resulting in

$$Y_{it} = \left(\frac{\mu_{it} \alpha \phi e^{\frac{\omega_{it} + \eta_{it}}{\alpha}}}{P_{Lit}} \right)^{\frac{\alpha}{1 - \phi\alpha}}$$

$$P_{it} = \mu_{it}^{\frac{1-\alpha}{1-\phi\alpha}} e^{(\omega_{it}+\eta_{it}) \frac{\phi - 1}{1 - \phi\alpha}} \left(\frac{\alpha\phi}{P_{Lit}} \right)^{\frac{\alpha(\phi - 1)}{1 - \phi\alpha}}$$

and profit-maximizing labor demand is

$$L_{it} = \left(\frac{\mu_{it} \alpha \phi}{P_{Lit}} \right)^{\frac{1}{1-\phi\alpha}} e^{(\omega_{it}+\eta_{it})\left(\frac{\phi}{1-\phi\alpha}\right)}$$

As expected, output price is negatively correlated with productivity and positively correlated with the demand shifter, while output is positively correlated with both shocks.

Taking logs we get

$$l_{it} = \frac{1}{1 - \phi\alpha} \ln(\phi\alpha) + \left(\frac{\phi}{1 - \phi\alpha} \right) (\omega_{it} + \eta_{it}) + \frac{1}{1 - \phi\alpha} \ln \mu_{it} - \frac{1}{1 - \phi\alpha} \ln P_{Lit}$$

and the estimated revenue function is

$$r_{it} = y_{it} + p_{it} = \alpha l_{it} + \omega_{it} + \eta_{it} + p_{it}$$

Because l_{it} is correlated with p_{it}, even after accounting for ω_{it} and η_{it}, omitting the output price from this regression results in an omitted variable bias. Using price changes to partially deflate revenues results in the estimated equation

$$r^*_{it} = r_{it} - \sum_{s=b+1}^{t} \Delta p_{is} = \alpha l_{it} + p_{ib} + \omega_{it} + \eta_{it}$$

As in Section 3.1 using price changes only does not solve the problem if the demand shifter μ_{it} is serially correlated because this serial correlation implies that p_{ib} is correlated with p_{it} and therefore with current labor input l_{it}.

The special case in Section 3.1 obtains when prices are exogenous, i.e., when $\phi = 1$. In this case, $P_{it} = \mu_{it}$ and

$$l_{it} = \frac{1}{1-\alpha}\ln\alpha + \left(\frac{1}{1-\alpha}\right)(\omega_{it} + \eta_{it}) + \frac{1}{1-\alpha}p_{it} - \frac{1}{1-\alpha}p_{Lit}$$

which is the case analyzed in Section 3.1 except that there η_{it} is observed after the labor choice is made.

B Simulation of model in Section 3.1

We assume the processes for (log) prices and productivity are AR(1),

$$\begin{aligned} p_0 &= \varepsilon_0 \qquad p_t = \rho p_{t-1} + \varepsilon_t \text{for} t \geq 1 \\ p_{L0} &= \varepsilon_{L0} \qquad p_{Lt} = \rho_L p_{Lt-1} + \varepsilon_{Lt} \text{for} t \geq 1 \\ \omega_0 &= \varepsilon_{\mu 0} \qquad \omega_t = \rho_\omega \omega_{t-1} + \varepsilon_{\omega t} \text{for} t \geq 1 \end{aligned}$$

where the firm i index is omitted and p and p_L are log output price and log wages, respectively.

All innovations to these processes are independent of each other and are drawn from a $N(0,1)$ distribution. This is a simplifying assumption implying no relationship between prices and productivity. The noise to the production function η_{it} is also $N(0,1)$ and independent of the other innovations.

We use the three exogenous processes for prices and productivity to generate data for labor using (6) and, adding the random noise η_{it} and ω_{it}, we generate log output using (5). We set $\alpha = 0.35$.

We simulate the model for 500 firms and let the simulations run over 100 periods. We discard the first 90 periods to be left with a balanced panel of 500 firms over 10 periods. We estimate parameters in each simulated sample. This is repeated 1000 times and the estimated parameters are averaged over the repetitions. These averages and the standard deviation of the estimated parameters across the 1000 repetitions appear in Table 1.

C The ESEE dataset

The ESSE (*Encuesta Sobre Estrategias Empresariales*) is an annual firm-level survey sponsored by the Spanish Ministry of Industry. The ESEE includes the balance sheet data and information on firm strategies of a representative sample of Spanish manufacturing firms. At the beginning of the survey in 1990, 5% of firms with up to 200 workers were randomly sampled by industry and size strata. All firms with more than 200 workers were asked to participate, which they did at a rate of about 70%. In order to preserve representativeness, samples of newly created firms were added to the initial sample every year to replace non-responding and exiting firms. In this illustrative application we use the subsample of small firms that entered the sample between 1991 and 1996 and stayed in the panel for at least four consecutive years. We end the sample period in 1999.

In what follows we define the variables that we use in the empirical application:

- *Revenue*. Value of produced goods and services, computed as sales plus the variation of inventories.
- *Capital*. Capital at current replacement values K_{it} is computed recursively from an initial estimate and using data on current investments in equipment goods I_{it}. We update the value of the past stock of capital by means of the price index of investment in equipment goods p_{It} as $K_{it} = (1-\delta)(p_{It}/p_{It-1})K_{it-1} + I_{it}$, where δ is an industry-specific estimate of the rate of depreciation. Capital in real terms is obtained by deflating capital at current replacement values by the price index of investment in equipment goods.
- *Employment*. Total hours worked computed as the number of workers times the average hours per worker, where the latter is computed as normal hours plus average overtime minus average working time lost at the workplace.
- *Materials*. Value of intermediate consumption (including raw materials, components, energy, and services)
- *Materials price index*: Firm-specific price index for intermediate consumption: firms are asked about the price changes that occurred during the year for raw materials, components, energy, and services. The price index is computed as a Paasche-type index.

- *Output price index*: Firm-level price index for output. Firms are asked about price changes (and the share) in up to five markets according ”a product line, type of consumers or another characteristic”. The exact wording of the question in the survey is: “State if the firm has changed, with respect to previous year, the price of the products sold in this market and the average change in percentual points”. The box of the question includes a separate space for the yes/not answer and the sign of the change and the percentage of change. The firm-level price index is computed as a Paasche-type index.