The inconsistency of De Loecker and Warzynski's (2012) method to estimate markups and some robust alternatives*

Ulrich Doraszelski
University of Pennsylvania†

Jordi Jaumandreu
Boston University‡

July 2020
Preliminary

Abstract

De Loecker and Warzynski (2012) estimate markups by computing the ratio of the estimated elasticity of a variable input to its (corrected) input share in revenue. The method overlooks that to consistently estimate input elasticities one needs to observe and control for marginal cost. We analyze the resulting biases that affect elasticities and markups and provide some robust alternatives.

*We are grateful to Ivan Fernandez-Val for useful dicussions, and to Rigoberto Lopez for comments.

†Wharton School, Email: doraszelski@wharton.upenn.edu

‡Department of Economics, Email: jordij@bu.edu

1. Introduction

An industry has gone through a wave of acquisitions and a researcher is asked by a policymaker to assess if the acquiring firms have increased their market power. She is given a panel data set including detailed firm-level production data, but neither the identity of the acquiring firms nor the moment of the acquisition are revealed. The researcher computes markups using the proposal of De Loecker and Warzynski (2012), henceforth DLW, and gives them to the policymaker to run a regression on a set of "step" dummies identifying the acquisitions and the "after" period. If the acquisitions have in truth raised markups, the regression is going to tell exactly the opposite. Below we explain what has gone wrong in the exercise of the researcher and how to solve the problem.

Are the firms resulting from merger and acquisitions decreasing or increasing their markups with respect to their pre-merger ones? Do exporters of similar products get larger or smaller markups than domestic sellers? Do trade liberalizations and tariff reductions enlarge or reduce market power of the firms? Does monopsony or bargaining in the labor market affect significantly the markups charged by firms? Are markups raising in developed economies? Economists would like to be able to give answers to these and similar questions with markups measurements that avoid making market behavior and technological assumptions. The so-called production function approach to measuring markups wants to develop ways to do so.[1] Moreover we would ideally like to measure markups without the need to know the precise agents and moments in which the changes happen, as the researcher of the opening paragraph has to do.

[1]See, for a recent panorama on measuring markups, Berry, Gaynor and Morton, 2019. Here we challenge, however, the view that the production function research "..provides presuasive evidence that markups have been rising..." (page 49). We also think of the production function approach as something usable at any level of aggregation at which it is applied.

The DLW proposal has been repeatedly cited and applied (see, for example, De Loecker, Goldberg, Khandelval and Pavcnik, 2016; Brandt, Van Biesebroeck, Wang and Zhang, 2017, 2019; De Loecker and Scott, 2016; De Loecker, Eeckhout and Unger, 2020; De Loecker and Eeckhout, 2018, and Author, Dorn, Katz, Patterson and Van Reenen, 2020). One reason is its simplicity: since the markup equals the ratio of the elasticity of a variable input to its (corrected) input share in revenue, it looks like it can be computed by plugging a consistent estimate of the input elasticity and of the disturbance of the production function that adjusts actual to unobservable output (and hence corrects the input share in revenue).

However, the application of DLW is producing a list of empirical findings at odds with predictions of economic theory and with alternative evidence.[2] Firms that export are found to get greater markups in the more competitive exports markets (De Loecker and Warzynski, 2012; see Doraszelski and Jaumandreu, 2019, for the reversal of the result with the same data).[3] Firms that face falling trade barriers and hence more competition are found to increase their markups and hence market power (De Loecker, Goldberg, Khandelval and Pavcnik, 2016; Brandt, Van Biesebroeck, Wang and Zhang, 2017, 2019). Firms subject to efficiency shocks and intense competition are found to experience steady increases in their market power (De Loecker, Eeckhout and Unger, 2020, and Author, Dorn, Katz, Patterson and Van Reenen, 2020; see Demirer, 2020, for an alternative measurement which halves the increase in market power accounting for flexibility in the production function and the presence of labor augmenting productivity).

We argue that the DLW method is inconsistent and hence produces biased markups, what can potentially explain a good deal of these bewildering results. There are two

[2] In addition, Raval (2020) shows that the DLW estimator produces systematically different results when applied with labor and materials.

[3] See Blum, Claro, Horstmann and Rivers, 2018, and Jaumandreu and Yin, 2019, for evidence on smaller margins of the exporters.

reasons. First, the researcher picks arbitrarily the functional form to estimate the elasticity without being advised any check about its adequacy. Second, and more important, DLW direct the researcher to estimate the elasticity of the input in the production function using a mispecified cost minimization relationship that omits the markup. This omission has two consequences. On the one hand, plugging the estimated disturbance into the formula to form the estimated markup induces inverse correlation with any true markup determinant that is unknown or has been left uncontrolled (the result of the exercise in the opening paragraph). On the other, the estimated elasticity is going to be biased by the use of instruments that are correlated with the uncontrolled part of the markup.

Under cost minimization, the elasticity of a variable input equals the share of this input in variable cost multiplied by the short-run elasticity of scale. It follows that DLW (without observational error) is theoretically equivalent to estimate the markup by the ratio revenue to variable cost (Bain 1951 classical solution) multiplied by the elasticity of scale. When the researcher chooses the production function is determining how the markup departs from the ratio price-average variable cost. For example, a constant gap if the production function is Cobb-Douglas, some variation with the logs of the variable inputs if it is a translog separable in K, and so on. The choice determines heavily the result and no test is performed of the agreement with the data.

On the other hand, the DLW method uses an Olley and Pakes (1996)/Levinsohn and Petrin (2003) procedure, henceforth OP/LP, implemented in the form proposed by Ackerberg, Caves and Frazer (2015), henceforth ACF, to estimate the input elasticities from the production function. These procedures, however, are ill-suited with markups varying across firms. In this note we show that to estimate the elasticity of a variable input by OP/LP when only cost minimization is assumed, it is necessary to observe marginal cost and hence the markup. This has also been recently argued by

Bond, Hashemi, Kaplan and Zoch (2020). The DLW method suffers from circularity: to estimate consistently an elasticity and the disturbance to be used in estimating the markup we would need to observe the markup. We analyze the biases incurred by ignoring the presence of the markup in the FOCs of the firm at the time to estimate the elasticity, and show that they are not small.

There is in addition a circumstance that makes the DLW proposal inconsistent both because functional form and method of estimation. It is the presence of non-neutral productivity. Researchers have recently stressed how labor-augmenting productivity in the production function may determine systematically varying elasticities, and explored ways to estimate it (Doraszelski and Jaumandreu, 2018, 2019; Raval, 2019; Demirer, 2020). On the one hand, this introduces an unobservable that affects directly the specification of the labor input and should be accounted for in the elasticity specification. On the other, it brings another unobservable to the procedures to estimate elasticities under unobserved productivity.

We close this note exploring some consistent alternatives to DLW that are robust in two senses: they do not depend on assumptions on the competitive behavior of firms, as DLW, and avoid any restrictive assumption on the form of the production function. We start by avoiding nesting a mispecified cost minimization relationship in the production function. At its simplest, this can be done by using dynamic panel methods to estimate the production function, although dynamic panel methods require stronger assumptions on the stochastic process for productivity. We also look, at the time to estimate the elasticities, at the FOCs and the production function as a system. With two variable inputs we can use one of the FOCs or the combination of the two.[4] The combined FOC, for example, tells us that the observed ratio revenue over variable cost (R/VC) is equal to the markup μ divided by the short-run elasticity

[4] If one of the inputs can be claimed without adjutment costs, it can be some advantage in doing what we propose using the FOC for this input. See Doraszelski and Jaumandreu (2019).

of scale ν or sum of the elasticities of the variable inputs up to an uncorrelated error. Assuming Hicks-neutral productivity, the elasticity of scale ν is a function of K, L and M only.[5] Estimating the markup up to the uncorrelated error implies to estimate this elasticity of scale (and use it to correct what we observe). We discuss how this estimation of the elasticity of scale can be done by means of the production function alone or by means of the FOC alone, as well as how the equations can be combined to obtain more efficiency.

Our exploration and empirical application raise a clear set of lessons. First, in estimating markups it is important to avoid restrictive functional forms for the production function. Second, estimation of the disturbance of the production function should be given up, and inference be based on the average of this disturbance across large enough groups of firms. Third, the combined FOC, with the ratio revenue to variable cost as dependent variable, is an excellent alternative relationship to the production function to estimate the elasticity of scale. Four, the FOC can be transformed, as it is usual to do with the production function, assuming that the unobserved markup follows an autoregressive process. This relaxes the moment assumptions needed to identify. Five, with the FOC differentiated or not, the instruments for K, L and M should be carefully chosen and checked by means of specification tests. Six, production function and FOC together can give more precise estimates.

The rest of this note is organized as follows. In Section 2 we recall the setup and the OP/LP procedure to estimate input elasticities. In Section 3 we show that, if marginal cost is not observed, an OP/LP procedure doesn't succeed in replacing unobservable productivity by observables. We also show that this cannot be solved by modeling unobservable marginal cost by means of observable average variable cost. Then, in Section 4 we analyze what happens when the researcher runs anyway a first stage ACF regression of output on inputs, input prices, and demand variables. We

[5] We later discuss how to deal with labor-augmenting productivity.

detail how is possible the result of the researcher of the opening paragraph if the markups have not fallen. An example helps to form a rough idea of the possible size of the biases in the estimation of elasticities too. Section 5 explores robust alternatives, organized in single-equation estimates (from the production function and from the FOC) and system of equations estimation. We illustrate the results with ten manufacturing unbalanced panel samples and briefly show the undesirable characteristics of DLW. Section 6 briefly comments three extensions: the dealing with non-neutral productivity, and two possible consistent solutions adding more structure to the system. Section 7 concludes.

2. Estimating markups

The setup

Because firms compete in a way that is unknown to the researcher, DLW only assume cost minimization. Firm j produces output Q_{jt} in period t with a given amount of capital, K_{jt}, and freely variable amounts of labor and materials, L_{jt} and M_{jt}, with production function

$$Q_{jt} = Q_{jt}^{*} \exp(\varepsilon_{jt}) = F(K_{jt}, L_{jt}, M_{jt}) \exp(\omega_{jt} + \varepsilon_{jt}),$$

where ω_{jt} is Hicks-neutral productivity that is observed by the firm but not by the econometrician. As usual in OP/LP, DLW assume that ω_{jt} follows a first-order Markov process with law of motion $\omega_{jt} = E(\omega_{jt}|\omega_{jt-1}) + \xi_{jt}$. The disturbance ε_{jt} is uncorrelated over time and with the inputs and accounts for the difference between unobserved planned output q_{jt}^{*} (we adopt lowercase letters to represent logarithms) and observed actual output $q_{jt} = q_{jt}^{*} + \varepsilon_{jt}$.

The firm minimizes variable cost $VC_{jt} = W_{jt}L_{jt} + P_{Mjt}M_{jt}$, where W_{jt} and P_{Mjt} are the price of labor and materials respectively. The variable input quantities are chosen to produce the planned quantity Q^*_{jt}. A slight rewriting of the cost minimization FOCs gives

$$\frac{1}{MC_{jt}(K_{jt}, W_{jt}, P_{Mjt}, Q^*_{jt}, \omega_{jt})} = \frac{\frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}} \exp(\omega_{jt})}{W_{Xjt}}, \tag{1}$$

for $X_{jt} = L_{jt}, M_{jt}$ and where W_{Xjt} is correspondingly W_{jt} or P_{Mjt}. The conditions say that the marginal productivity of the last dollar must be equal in every use, as Samuelson (1947) puts it and Hall (1988) first empirically exploited.

DLW estimate the markup $\mu_{jt} = \frac{P_{jt}}{MC_{jt}}$ from equation (1) by multiplying both sides by P_{jt}, completing conveniently numerator and denominator, and replacing the variable input elasticity $\beta_{Xjt} = \frac{X_{jt}}{Q^*_{jt}} \frac{\partial Q^*_{jt}}{\partial X_{jt}}$ and the disturbance ε_{jt} by estimates. That is

$$\widehat{\mu}_{jt} = \frac{\widehat{\beta}_{Xjt}}{S^R_{Xjt}} \exp(-\widehat{\varepsilon}_{jt}), \tag{2}$$

where $S^R_{Xjt} = \frac{W_{Xjt} X_{jt}}{P_{jt} Q_{jt}}$ is the input share in revenue. Note that S^R_{Xjt} is based on actual output Q_{jt}.

Using an OP/LP procedure to estimate an input elasticity

DLW estimate the input elasticity β_{Xjt} and production function disturbance ε_{jt} using an OP/LP procedure, implemented as suggested by ACF. Any OP/LP procedure starts by looking for a relationship to invert in order to replace ω_{jt} by observables in

$$q_{jt} = q^*_{jt} + \varepsilon_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + \omega_{jt} + \varepsilon_{jt}. \tag{3}$$

OP use the demand for investment, LP the demand for a variable input. The researcher postulates a function $\omega_{jt} = h(x_{jt})$, where x_{jt} collects input quantities and all arguments of the input demand that is inverted to control for ω_{jt}, including input prices. When plugging $h(x_{jt})$ in (3) the expression can be written as

$$q_{jt} = \phi(x_{jt}) + \varepsilon_{jt},$$

where $\phi(\cdot)$ is an unknown functional form.

This allows, in a first step, to carry out a nonparametric regression and estimate the disturbance ε_{jt} that separates observed output from planned output $q^*_{jt} = \phi(x_{jt})$. In a second step, the production function is estimated in the regression

$$q_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + g[\widehat{\phi}_{jt-1} - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})] + \xi_{jt} + \varepsilon_{jt},$$

based on the Markovian assumption on productivity, where $g(\cdot)$ is an unknown functional form. Note that the inputs chosen at moment t, after the realization of ξ_{jt}, will be correlated with this innovation and need to be intrumented.

The DLW method includes more variables in the postulated function, besides input quantities and prices, and thinks of it as

$$\omega_{jt} = h(x_{jt}, z_{jt}),$$

where z_{jt} consists of "additional variables potentially affecting difference in input demand choices of firms" (De Loecker and Warzynski (2012), page 2446). We correspondingly enlarge the function $\phi(\cdot)$ writing $q_{jt} = \phi(x_{jt}, z_{jt}) + \varepsilon_{jt}$.

3. Inverting for unobserved productivity

Inverting variable inputs

The variable cost function is $VC(K_{jt}, W_{jt}, P_{Mjt}, Q^*_{jt}/\exp(\omega_{jt}))$. Therefore $MC_{jt} = VC_4(K_{jt}, W_{jt}, P_{Mjt}, Q^*_{jt}/\exp(\omega_{jt}))\exp(-\omega_{jt})$, where VC_4 denotes the derivative of variable cost with respect to its fourth argument.[6] Plugging into (1) the terms $\exp(\omega_{jt})$ in both sides cancel and we have

$$\frac{1}{VC_4(K_{jt}, W_{jt}, P_{Mjt}, Q^*_{jt}/\exp(\omega_{jt}))} = \frac{\frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}}}{W_{Xjt}}.$$

[6] With labor predetermined, $MC_{jt} = VC_4(K_{jt}, L_{jt}, P_{Mjt}, Q^*_{jt}/\exp(\omega_{jt}))\exp(-\omega_{jt})$. Inverting materials, for example, produces what ACF call a materials input demand conditional on labor.

This can be inverted to get

$$\frac{Q_{jt}^*}{\exp(\omega_{jt})} = H(K_{jt}, L_{jt}, M_{jt}, W_{jt}, P_{Mjt})$$

but, because Q_{jt}^* is unobservable, not to get a function $\omega_{jt} = h(x_{jt})$ that expresses ω_{jt} as a function of observables. This observation generalizes if we invert a combination of FOCs rather than a single FOC.

Alternatively, using Shephard's Lemma, the input demands are

$$X_{jt} = VC_X(K_{jt}, W_{jt}, P_{Mjt}, Q_{jt}^*/\exp(\omega_{jt})),$$

where VC_X denotes the derivative of variable cost with respect to the price of input X_{jt}. Again, these can be inverted for $\frac{Q_{jt}^*}{\exp(\omega_{jt})}$ but not for ω_{jt} as a function of observables.

Inverting investment

The demand for investment is usually assumed to come from dynamic profit maximization, and thus requires the researcher to take a stand on competition and demand, what DLW want to avoid. One may, alternatively, suppose that underlying the demand for investment is a dynamic cost minimization problem (see, for example, Doraszelski and Jaumandreu (2019)). In this case investment will be a function of the ratio $Q_{jt}^*/\exp(\omega_{jt})$ and an analogous problem to the one just reviewed arises.

Replacing unobserved MC by observed AVC

If we observe MC_{jt}, then the FOCs would provide an expression for ω_{jt} in terms of observables as (1) makes clear.[7] But then we also could directly compute the markup $\mu_{jt} = \frac{P_{jt}}{MC_{jt}}$ without first estimating input elasticities from the production function.

[7] Under profit maximization Doraszelski and Jaumandreu (2013) invert the equivalent expression in terms of marginal revenue, that then is estimated simultaneously with the production function.

One may wonder what happens if one relies on the observable average variable cost $AVC_{jt} = \frac{VC_{jt}}{Q_{jt}}$ to replace marginal cost by

$$MC_{jt} = \frac{VC_{jt}}{\nu_{jt} Q^*_{jt}} = \frac{AVC_{jt}}{\nu_{jt}} \exp(\varepsilon_{jt}),$$

where $\nu_{jt} = \beta_{Ljt} + \beta_{Mjt}$ is the short-run elasticity of scale. Plugging this expression into (1) we get

$$\exp(\omega_{jt}) = \frac{W_{Xjt}}{\frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}}} \frac{\nu_{jt}}{AVC_{jt}} \exp(-\varepsilon_{jt}).$$

When the log of this expression is substituted for ω_{jt} in equation (3), we get

$$q_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + w_{Xjt} - \ln \frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}} + \ln \nu_{jt} - avc_{jt},$$

or

$$\beta_{Xjt} = \nu_{jt} S_{Xjt},$$

where $S_{Xjt} = \frac{W_{Xjt} X_{jt}}{VC_{jt}}$ is the share of the input in variable cost. This is not useful, since the aim of the first step of ACF is estimating the disturbance ε_{jt} which drops from the relationship.[8]

4. Estimating with bias

As shown in the previous section, when we do not observe MC_{jt} there does not exist a function $\omega_{jt} = h(x_{jt})$ that allows replacing productivity ω_{jt} by observables x_{jt} because planned output Q^*_{jt} is unobservable. The intuition behind DLW inclusion of z_{jt} in $h(\cdot)$ is that, under imperfect competition, a firm's planned output q^*_{jt} has a structural demand explanation that brings variables different from x_{jt}.[9] Let's write

[8]The expression for ω_{jt} in terms of AVC_{jt} can however be used lagged, together with a linear Markovian specification, to construct an estimable model.

[9]Strictly speaking DLW include in z_{jt} some variables, such as input prices, that should already have been included in x_{jt}. A good example of other variables in z_{jt} is De Loecker, Goldberg, Khandelval and Pavcnik (2016): location, product dummies, export status, input and output tariffs, market share and ouput price.

this demand as $q^*_{jt} = \ln D(z_{jt}) + \delta_{jt}$, where the unobservable δ_{jt} represents demand heterogeneity. The only possibility to have a function only of observables is that $\delta_{jt} = 0$, so we need a complete knowledge of the demand of the firms. This is unlikely given the role that the literature has uncovered for demand heterogeneity. Firm-level demand estimation in the tradition of Berry, Levinsohn and Pakes (1995) stresses the importance of the remaining "unknown characteristic" even after including observed product attributes. Estimation without detailed characteristics has repeatedly found that unobserved demand heterogeneity explains the same or more output variation than unobserved productivity.[10]

What happens when the researcher performs a nonparametric regression of q_{jt} on x_{jt} and z_{jt} in the first stage of ACF? The researcher gets the conditional expectation of q_{jt} on x_{jt} and z_{jt} that, if ε_{jt} is uncorrelated with all the regressors, is equal to the conditional expectation of q^*_{jt} on x_{jt} and z_{jt} plus a residual r_{jt} in addition to the disturbance ε_{jt} :

$$q_{jt} = \widetilde{\phi}(x_{jt}, z_{jt}) + r_{jt} + \varepsilon_{jt},$$

where $r_{jt} = q^*_{jt} - E(q^*_{jt}|x_{jt}, z_{jt})$ Notice that r_{jt} is by construction mean-independent of x_{jt} and z_{jt} and hence $\widetilde{\varepsilon}_{jt} = r_{jt} + \varepsilon_{jt}$ is uncorrelated with x_{jt} and z_{jt}.

The conditional expectation of q^*_{jt} can be computed at least in two ways, as $E(q^*_{jt}|x_{jt}, z_{jt}) = \ln F(K_{jt}, L_{jt}, M_{jt}) + E(\omega_{jt}|x_{jt}, z_{jt})$ and as $E(q^*_{jt}|x_{jt}, z_{jt}) = \ln D(z_{jt}) + E(\delta_{jt}|x_{jt}, z_{jt})$, so the residual can be written at least in two ways too

$$r_{jt} = \omega_{jt} - E(\omega_{jt}|x_{jt}, z_{jt}) = \delta_{jt} - E(\delta_{jt}|x_{jt}, z_{jt}).$$

The unobservability associated with all variables that covary with q^*_{jt} but that are not included in x_{jt} or z_{jt} remains in r_{jt}. Later we show that this renders the second stage of ACF inconsistent.

[10] See, for recent assessments, Eslava and Haltiwanger (2020) and Jaumandreu and Yin (2020).

However, there is still something more. Unobservable productivity is in fact closely related to the markup given the observed variables. The relationship is

$$\begin{aligned}\ln\mu_{jt} &= \ln P_{jt} - \ln MC(K_{jt}, W_{jt}, P_{M_{jt}}, Q^*_{jt}, \exp(\omega_{jt})) \\ &= -\ln\overline{MC}(K_{jt}, M_{jt}, \frac{W_{jt}}{P_{jt}}, \frac{P_{Mjt}}{P_{jt}}) + \omega_{jt}. \end{aligned} \tag{4}$$

The first equality comes directly from the definition of μ_{jt}. The second uses the replacement of $Q^*_{jt}/\exp(\omega_{jt})$ by M_{jt} in the derivative of $VC(\cdot)$ with respect to Q^*_{jt} (page 9), and then applies the linear homogeneity of $MC(\cdot)$ in input prices to rewrite the observable part of marginal cost $\overline{MC}(\cdot)$ in terms of real prices. This implies that equation (3) can also be written as

$$q_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + \ln\overline{MC}(K_{jt}, M_{jt}, \frac{W_{jt}}{P_{jt}}, \frac{P_{Mjt}}{P_{jt}}) + \ln\mu_{jt} + \varepsilon_{jt}.$$

This is telling us another way to read the residual r_{jt}. It is going to convey all the determinants of the markup that are not controlled for in vector (x_{jt}, z_{jt}) :

$$r_{jt} = \ln\mu_{jt} - E(\ln\mu_{jt}|x_{jt}, z_{jt}).$$

We are first going to see how this biases the markups.

The bias generated by the estimated residual

The fists step of DLW gets an estimate of $\widetilde{\varepsilon} = r + \varepsilon$, and this estimate is used in the formula for the computation of the markup as $\exp(-\widehat{\widetilde{\varepsilon}})$. From the fact that $r_{jt} = \ln\mu_{jt} - E(\ln\mu_{jt}|x_{jt}, z_{jt})$ it follows that the DLW estimate of the markup is going to be negatively correlated with any true determinant of the markup that has not been controlled for.

If the researcher has included input quantities and prices, three leading examples of variables possibly remaining in r_{jt} are the heterogeneity and shocks of demand as well as any surprises that can have affected the markup and, by its character, are scarcely or not correlated with the included variables. It is likely that any uncontrolled labor-augmenting productivity is in r_{jt} too.

The bias in the estimation of the elasticity

If $\delta_{jt} \neq 0$, the second stage of ACF is based on the regression

$$\begin{aligned} q_{jt} &= \ln F(K_{jt}, L_{jt}, M_{jt}) \\ &\quad + g[\widetilde{\phi}(x_{jt-1}, z_{t-1}) + r_{jt-1} - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})] + \xi_{jt} + \varepsilon_{jt}. \end{aligned} \tag{5}$$

As in many applied exercises, take $g(\cdot)$ as an $AR(1)$ process with parameter ρ.[11] Then the term ρr_{jt-1} becomes part of the composite error of the regression.

The estimation of (5) calls for moments based on instruments that can be assumed orthogonal to the innovation of the Markov process ξ_{jt}.[12] Variables K_{jt} and L_{jt} are routinely used as instruments (as De Loecker, Goldberg, Khandelval and Pavcnik (2016) do, invoking ACF). The assumption is that the firm chooses these variables at $t-1$ and are hence uncorrelated with ξ_{jt}. This timing assumption is prevalent for capital and corresponds to the possible assumption of fixed labor in the short-run considered by ACF. With the presence of r_{jt-1}, none of these instruments is valid. Any variable that has been chosen at $t-1$ has been chosen according to the the values

[11] Alternatively, we can approximate a nonlinear $g(\cdot)$ around the value $r_{jt-1} = 0$. The Taylor expansion has the form

$$\begin{aligned} q_{jt} &= \ln F(K_{jt}, L_{jt}, M_{jt}) + g(\widetilde{\phi}(x_{jt-1}, z_{jt-1}) - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})) \\ &\quad + g_s(\omega'_{jt-1} - r_{jt-1}) r_{jt-1} + \textstyle\sum_{s=2}^{\infty} g_s(\omega'_{jt-1} - r_{jt-1}) r^s_{jt-1}/s! + \xi_{jt} + \varepsilon_{jt}, \end{aligned}$$

where g_s denotes the $s-th$ derivative of $g(\cdot)$ and $\omega'_{jt-1} \equiv \widetilde{\phi}_{jt-1} + r_{jt-1} - \ln F_{jt-1}$. The remaining composite error is a function of the lagged value of the unobservable that the first step has tried to eliminate. There is no IV solution to the resulting problem.

[12] The problem usually calls for at least four moments, in addition to the one based on the constant, to identify four parameters (three elasticities and parameter ρ).

of ω_{jt-1} and δ_{jt-1} and is therefore correlated with what remains of ω_{jt-1} and δ_{jt-1} in r_{jt-1}.[13]

Variables $K_{jt-1}, L_{jt-1}, M_{jt-1}$ are valid as instruments because the firm decides on them before it observes ξ_{jt}. These variables are correlated with with ω_{jt-1} and δ_{jt-1} but, if they have been included in x_{jt-1}, are uncorrelated by construction with r_{jt-1}. For the same reason input prices and possible demand shifters dated at $t-1$ that have been included in x_{jt-1} and z_{jt-1} are uncorrelated with r_{jt-1}. Input prices and shifters, however, may require an specific justification as instruments because they are variables potentially correlated with the innovation of the Markov process ξ_{jt}.[14] Lagged output price if included in z_{jt-1} can be an instrument.

The sign and size of the elasticity bias

We work out the sign and size of the bias due to using L_{jt} as instrument considering a simple example of a single-input (labor) production function: $q = \beta l + \omega + \varepsilon = q^* + \varepsilon$. We assume that labor L_{jt} is chosen at $t-1$, as ACF and De Loecker, Goldberg, Khandelval and Pavcnik (2016) do. The demand function is $q^* = -\eta p + \delta$, where p is output price and η the elasticity of demand and δ is what remains unexplained of q^* from the point of view of demand after controlling for the output price. Our x is now l, our z is now p. We consider population regressions and drop firm and time subscripts for simplicity.

The researcher observes q, l, and p, and regresses q on l and p to predict q^* in the

[13] Eslava and Haltiwanger (2020) is an example of problematic identification. Their moment conditions (22) not only use contemporaneous K and L but assume that lagged demand heterogeneity is orthogonal to the production function disturbance that is left after applying the first stage regression. We just have shown that, with DLW, the first stage ACF regression leaves demand heterogeneity partly uncontrolled in the production function equation.

[14] The increase of the price paid for an input, for example, may be correlated with a subsequent productivity increase.

first stage of ACF. The result is $q^* = \widetilde{\phi}(l,p) + r$ where, according to our previous explanation, $r = \omega - E(\omega|l,p) = \delta - E(\delta|l,p)$.

The second stage of ACF takes the form of the regression $q = \beta l + \rho(\widetilde{\phi}(l_{-1}, p_{-1}) + r_{-1} - \beta l_{-1}) + \xi + \varepsilon$. Assume that the $AR(1)$ parameter ρ is known. Using the knowledge of ρ and $\widetilde{\phi}(l_{-1}, p_{-1})$, the relevant equation to determine the bias becomes $\overline{q} = q - \rho\widetilde{\phi}(l_{-1}, p_{-1}) = \beta(l - \rho l_{-1}) + \rho r_{-1} + \xi + \varepsilon$. The IV estimator for β, using l as instrument, is

$$\beta_{IV} = \frac{E(l\overline{q})}{E(l(l-\rho l_{-1}))} = \beta + \frac{\rho E(l r_{-1})}{E(l(l-\rho l_{-1}))}.$$

The bias is determined by the correlation between l and $r_{jt} = \omega_{-1} - E(\omega_{-1}|l_{-1}, p_{-1})$. In general, the correlation between the labor input and productivity is considered to be positive (the output effect dominates the productivity effect), so we expect a positive bias. Another way to see the bias is to think of $r_{jt} = \delta_{-1} - E(\delta_{-1}|l_{-1}, p_{-1})$. Demand heterogeneity is likely to be positively correlated with the quantity of labor.

To work out the sign and size of the bias, the IV estimator can be written (see the Appendix) as

$$\beta_{IV} = \beta \left(1 + \rho \frac{1}{\sqrt{Corr(\widetilde{l}, r_{-1})^2 + \frac{R^2}{(1-R^2)}} - Corr(\widetilde{l}, r_{-1})} \frac{Corr(l, r_{-1})}{Corr(l, \widetilde{l})} \right), \quad (6)$$

where $\widetilde{l} = l - \rho l_{-1}$, and $R^2 = 1 - \frac{Var(r)}{Var(q^*)}$ is the proportion of the variance of q^* that is accounted for the explanatory variables in the population regression.[15] The formula reveals that, even if r_{-1} is a very small proportion of the variance so that R^2 is close to 1, the key for the quantitative importance of the bias is the value of the correlation between the instrument and the residual relative to the correlation between the instrument and the explanatory variable of the regression.[16]

[15] The formula assumes a stationary environment and, without loss of generality, that the instrument is in differences with respect to its mean.

[16] In practice the researcher will get the R^2 corresponding to the residual $r_{jt} + \varepsilon_{jt}$ in the regression of q_{jt}, but we do not take any stance on residual ε_{jt}.

The following table assumes that the $AR(1)$ parameter is 0.90 and that l follows an $AR(1)$ process with parameter 0.6. We consider three degrees of correlation of the instrument l with the residual r_{-1}, which according to equation (6) enters divided by $Corr(l, l - \rho l_{-1}) = 0.538$. The researcher who estimates the (log) markup using equation (2) gets

$$\ln \widehat{\mu} = \ln \beta + \ln(1 + bias) - \ln S^R - \widetilde{\varepsilon} \simeq \ln \mu + bias - r.$$

Biases of DLW markups

		R^2		
		0.99	0.95	0.90
	0.30	0.052	0.124	0.186
$Corr(l, r_{-1})$	0.45	0.079	0.192	0.293
	0.60	0.107	0.266	0.412

Assuming that markups are in the range of 10 to 20 percentage points, the implied percentage points of bias can go from significant (5 percentage points) to overwhelming (more than 40 percentage points).[17]

An interesting case appears when we have more than one input. The bias may be favoring some inputs against others, for example materials against labor since materials is likely to be more flexible in the short run.[18] With varying elasticities the bias will vary across firms and over time. Notice that is possible that the biases in the elasticity and the estimation of the residual go in different directions making unpredictable the total effects.

[17] The computed biases increase if the Markov process is less persistent or the labor process is more persistent.

[18] For example, Brandt, Van Biesebroeck, Wang and Zhang (2017) use the DLW estimator with the same data as Jaumandreu and Yin (2020), who control for productivity and demand heterogeneity. The first paper estimates a quite close average short-run scale parameter to the estimate of the second (0.96 versus 0.90) but a completely different split into elasticity of labor and materials: 0.05 and 0.91 versus 0.29 and 0.61.

5. Robust consistent alternatives

We present consistent alternatives that are robust in two senses. First, we do not need to specify any kind of behavior, we only assume cost minimization. Second, in clear progress relative to the estimators proposed up to date we do not need to specify any functional form for the production function. We base the estimates on a fully nonparametric production function.

We discuss the alternatives and their application dropping firm and time subindices to simplify the notation.

Estimating the short-run elasticity of scale

The error of DLW lies in taking OP/LP as an approximate method that frees the researcher of the discussion of what is exactly needed in order to write unobserved productivity as a function of observables. In fact, in this review has become evident that OP/LP is not able to deal with unspecified heterogeneous behavior on the part of the firms because heterogeneous markups break the "scalar unobservable assumption" in the FOCs.[19]

However it is true that cost minimization, properly used, brings additional information, contained in the first order conditions, to estimate the markup. Rewrite the formula of DLW for the FOCs considering the objects to be estimated as the right hand side of the variable that is observed. Taking the two variable inputs simultaneously we have

$$\frac{R}{VC} = \frac{\mu}{\nu} e^{\varepsilon},$$

where revenue R over variable costs VC is the inverse of the joint share in revenue of

[19] It continues of course being a nice alternative if the reseacher is willing to specify the behavior, and hence the markups, and estimates them simultaneously.

variable inputs, and recall that $\nu = \beta_L + \beta_M$ is the variable inputs joint elasticity.[20] This combined FOC shows that the problem consists of separating the unobservables μ and ν from the unique observable $\frac{R}{VC}$. If productivity is exclusively Hicks-neutral, the elasticity of scale is a function of the inputs only, $\nu = \nu(K, L, M)$. We will later deal with the case in which productivity is not neutral. On the other hand ε is an uncorrelated error.[21] It follows that, in the FOC, to estimate the markup we need to estimate $\nu(K, L, M)$ controlling by $\ln \mu$ in

$$\ln \frac{R}{VC} = -\ln \nu(K, L, M) + \ln \mu + \varepsilon. \tag{7}$$

It is a problem formally similar to the estimation of productivity. To estimate ω the researcher needs to estimate $F(K, L, M)$ in

$$q = \ln F(K, L, M) + \omega + \varepsilon, \tag{8}$$

controlling by ω.

Notice that equations (7) and (8) are linked by the restriction $\ln \nu(K, L, M) = \ln(\frac{\partial \ln F}{\partial l} + \frac{\partial \ln F}{\partial m})$. At the same time, they are very different. The second is the primal relationship, the first belongs to the duality that emerges from the firm minimizing costs.

To complete the setting it is convenient to recall the link between $\ln \mu$ and ω, given the observed variables, as expressed by equation (4)

$$\ln \mu = -\ln \overline{MC}(K, M, \frac{W}{P}, \frac{P_M}{P}) + \omega.$$

[20] Under cost minimization $\beta_X = \nu S_X$, where S_X is the share in variable cost. It follows that the DLW formula for each input collapses to this common formula. With both inputs variable, there is no reason to prefer one FOC to the combination. If only one input can be assumed without ajustment costs, the FOC based on this input avoids the need to specify adjustment costs. The discussion that follows can then be set in terms of this FOC.

[21] If we add problems of observability in the price the error is enlarged but is likely to continue being uncorrelated.

This expression shows how easily we can rewrite both the FOC and the production function equations in terms of observable variables and the cross-unobservable, but also that we cannot eliminate both simultaneously (no OP/LP procedure is available).

Estimation of the elasticity of scale ν can in principle be done by means of the production production function alone or by means of the FOC alone. To estimate the elasticity of scale in the production function we need to control for unobserved productivity ω. To estimate the elasticity of scale from the ratio revenue to variable cost we need to control for unobservable markup μ. As no OP/LP procedure is available, in both cases the problem of estimation is how to obtain consistency using the right instruments for K, L and M combined with a possible transformation of the equation that makes them legitimate.

It should be noticed, however, that the FOC alone cannot separate the constant of the elasticity of scale and the constant of the log of the markup, while the production function allows to compute this constant as the sum of the derivatives of the linear terms of the effects of l and m. This sounds strange because we are used to see markups in levels, but in fact levels are not necessary if the goal is to assess their evolution over time or compare them across firms. On the other hand, adopting an external estimate for the mean $\ln \nu(K, L, M)$, that is typically very close to zero, seems a reasonable solution. For example, we can borrow the estimate for the mean of $\ln \nu(K, L, M)$ from the production function.

Estimating from the production function

Assuming that ω follows an $AR(1)$ process the production function can be quasidifferentiated, the terms ω and ω_{-1} disappear from the equation, only the random innovation of the process is left and, in this context, K, L_{-1} and M_{-1} are accepted with generality as valid instruments. This procedure to deal with the estimation of the production function has been called "dynamic panel estimation" (Blundell and

Bond, 2000)[22] If the true process is a nonlinear Markov process, this will provide only an approximation, and a residual containing ω_{-1} will remain. Notice that the modeling of ω as an exogenous Markov process can also neglect the shifts due to the determinants of productivity, that remaining in the residual may be an additional reason for bias of the coefficients of the production function.[23] All this makes very important testing the specification obtaining evidence on the validity of the instruments (see Arellano and Bond, 1991). We use this estimate as a first possible consistent alternative.

The estimate for $\ln \nu(K, L, M)$ is computed as $\ln(\frac{\partial \ln F}{\partial l} + \frac{\partial \ln F}{\partial m})$ after estimating the production function, and the markup is estimated as $\ln \frac{R}{VC} + \ln(\frac{\widehat{\partial \ln F}}{\partial l} + \frac{\widehat{\partial \ln F}}{\partial m})$. It holds that

$$E(\ln \frac{R}{VC} + \ln(\frac{\widehat{\partial \ln F}}{\partial l} + \frac{\widehat{\partial \ln F}}{\partial m})) = \ln \mu.$$

In the results part we label this estimator production function or PF.

Digression: what is the markup and moments involving the markup

We want to proceed similarly in an estimation based on the FOC. We need instruments for K, L and M that are uncorrelated with the markup. It is worthy to spend some time discussing what are the moments that we can assume involving the markup.

Expression (4) gives only a definitional description of the components involving the markup $\ln \mu$. The markup is an endogenous variable that the firm impacts indirectly

[22] See the comparison with OP/LP procedures in Ackerberg, Caves and Frazer (2015) and Ackerberg (2020).

[23] Doraszelski and Jaumandreu (2013) is an example of estimating endogenous productivity, with the Markov process shifting due to the $R\&D$ investments of the firms. A long list of studies has considered innovations, exports and inports, entry of rivals, and other variables as shifters of productivity.

by setting the price and making changes that affect marginal cost, or just directly choosing it as objective (Jaumandreu and Lin, 2018, suggest that much is gained by modeling the firms' dynamic setting of the markup). It is convenient to distinguish the long-run and the short-run, that is what we are trying to measure. In the long-run, industrial organization analysis has typically considered the markup as the result of the elasticity of demand and the competitive behavior of firms. In the short-run, however, the markup is continuously affected by three kind of shocks: changes in input prices, productivity, and demand shocks.

With respect to the long-run equilibrium we should take into account the likely relationship with the variations in ucK/R, where uc is the user cost of capital, or component of the cost of capital in the economic profit rate. Even if we want to be agnostic about how this affects the markup we should expect some relationship with uc or K/R, specially if they vary a lot across units or over time. Interestingly enough, this does not imply necessarily any relationship with K or its variation (K and K/R tend to be uncorrelated).

In the short-run, there are at least two motives to consider that the persistence of the markup, suggested by the persistence in price average cost margins, has a Markovian form. First because productivity ω impacts linearly the log of the markup given other observed variables (see expression (4)). Second, because if the firm reacts with dynamic optimality to the impacts of input prices growth, productivity changes and demand shocks, the markup is likely to follow a first order stochastic difference equation of the form $\ln \mu = h(\ln \mu_{-1}, shocks)$. Of course the possibility of nonlinearity is again a question and, in addition, the combination of more than one linear process can raise higher order $ARMA$ processes.

The implications for possible moments involving the markup are as follows. Correlation with capital may be simply nonexistent. And the correlation between the markup and variable inputs is likely to be indirect. The input price and demand shocks can generate correlation between the changes of the markups and changes of the variable inputs, but this correlation is unlikely to be with the levels of the inputs. This suggests two versions for possible moments. The first exploits the lack of correlation of capital and the lagged inputs with the log of the markup

$$E[(k, k_{-1}, l_{-1}, m_{-1}) \ln \mu] = 0,$$

the second, much weaker, assumes a linear Markovian process for the markup, $\ln \mu = \rho_\mu \ln \mu_{-1} + \zeta$, and the orthogonality of capital and the inputs lagged with the innovations in the markup

$$E[(k, k_{-1}, l_{-1}, m_{-1})\zeta] = 0.$$

Estimating from the first order condition

In this case we directly estimate

$$\ln \frac{R}{VC} = g_0 + g(K, L, M) + \ln \mu + \varepsilon.$$

We will estimate g_0 from equating the mean value of $g_0 + g(K, L, M)$ to the mean value for $-\ln \nu(K, L, M)$ obtained in the production function. The estimate of the markup is then $\ln \frac{R}{VC} - \widehat{g}_0 - \widehat{g}(K, L, M)$ and again we have

$$E(\ln \frac{R}{VC} - \widehat{g}_0 - \widehat{g}(K, L, M)) = \ln \mu.$$

As remarked before, we use the same set of instruments when we assume that $\ln \mu$ follows an $AR(1)$ and we quasidifferentiate the equation. If the $AR(1)$ is true eliminates from the equation the levels of $\ln \mu$ (except from the long run equilibrium) and this should weaken additionally any correlation with the levels of the instruments. We will see later that the probability values of the specification test are much higher after differentiating the equation.

The FOC in levels and the FOC quasidifferentiated provide alternative consistent estimates to PF. Notice that these estimates do not depend on any assumption on the unobservable ω. This makes them robust to the nonlinearity of ω or its endogeneity. But they depend on the assumptions about the correlations between the inputs and the markup $\ln \mu$. Testing the assumptions is again a key. In the results part, we call these estimators first order condition in levels or $FOCL$ and first order condition or FOC, respectively.

Estimating from the system

Since both the FOC and the production function can provide independently consistent estimators, it seems natural to try to increase the precision by estimating the system of equations subject to the cross-equation constraint of equality of the elasticity of scale. We apply this assuming that both $\ln \mu$ and ω follow autoregressive processes that allow quasidifferentiation. That is, we estimate the cross-restricted system

$$\begin{aligned} \ln \frac{R}{VC} &= -\ln\left(\frac{\partial \ln F}{\partial l} + \frac{\partial \ln F}{\partial m}\right) + \ln \mu + \varepsilon, \\ q &= \ln F(K, L, M) + \omega + \varepsilon, \end{aligned}$$

with $\ln \mu = \rho_\mu \ln \mu_{-1} + \zeta$ and $\omega = \rho_\omega \omega_{-1} + \xi$. Once estimation has been carried out, we again construct the variable whose expectation is the markup. A great advantage of this system is that provides naturally an estimate for a mean of the elasticities of scale separated from he constant of $\ln \mu$ in the FOC. But notice that the system makes the estimation to depend on the assumptions on ω. We call this estimate system estimate or SYS.

To check the robustness of the results we specify two models that, in addition to quasidifferentiation, include explicitly in the FOC equation the determinants of the shifts in the markups. We expect this to weaken any possible remaining correlation between the instruments and what is left in the error. We assume first that the markup follows the process $\ln \mu = s\alpha + \rho_\mu \ln \mu_{-1} + \zeta$, where s is a vector of shifters. We include in s the changes in capital, in the input prices and an indicator of the state of demand. We will call this estimator controlled system or CSYS. The other model uses equation (4) to fully specify the observable part of the marginal cost in terms of real input prices and a flexible form for K and M. We enter prices linearly and assume that the $AR(1)$ process controls for productivity. This last specification can only be identified in the framework of the system (now some coefficients of the FOC cannot be identified with the FOC alone). We call it system with marginal cost or $SYSMC$.

Results

We apply the robust consistent alternatives to estimate the markups of a breakdown of manufacturing in 10 firm-level industry unbalanced panel samples. The data come from the Spanish Encuesta Sobre Estrategias Empresariales (ESEE) from 1990-2012. Doraszelski and Jaumandreu (2019) provide details on the sample and variables.

Recall that the aim is correcting $\ln(R/VC)$ from the variation in the log of the ratio AVC/MC. First we estimate the distribution of elasticities of scale from the production function (PF estimator). Then we use exclusively the first order condition, first in levels and then taking quasidifferences ($FOCL$ and FOC estimators). Next we estimate the elasticity of scale simultaneously from the production function and the first order condition, imposing the cross-restriction that the effects of L and M in the first order condition are the derivatives of the production function (SYS estimator). We end the exercise by performing the two robustness checks mentioned in the previous subsection (estimators $CSYS$ and $SYSMC$). We add a few insights on the results of the DLW estimators applied with Cobb-Douglas and Translog production functions in the same samples.

Results of the robust consistent estimators

Table 2 shows the average markups, and their sample standard deviations, resulting from the alternatives PF, $FOCL$, FOC, SYS, $CSYS$ and $SYSMC$, and Table 3 the correlations among the markups estimated in each case. Notice that the means of the markups obtained from PF, $FOCL$, FOC are the same, because we are imposing the mean correction obtained with PF, but also that the standard deviations reveal very different distributions. Elasticities of scale are, except for the $FOCL$ estimate (the only one carried out in levels), winsorized at the 0.05 and 0.95 percentiles.[24] In the appendix tables under label A2 we provide very important detail to assess the estimators. First the atoregressive coefficients estimated when the model implies quasidifferentiation. Second, the average value and standard deviation of the estimated distributions of the elasticity of scale, as well as the proportion of negative markups that they imply. Third, the result of regressing the estimated markups on the indicator of the state of demand as a check.

We always include in estimation a set of time dummies, that we systematically attribute to the variation over time of productivity and the markup respectively. For the flexible specification of the production function we use a complete polynomial of order three in K, L and M (19 terms). The flexible function $g(\cdot)$, collecting the effects of K, L and M in the FOC, is also approximated with a polynomial of order three. When we apply SYS, the polynomial in the FOC is restricted to the terms of the sum of the derivatives of the production function with respect to m and l.[25]

[24] In general this eliminates some extreme values affecting very little to the mean and dispersion of the distribution. We provide more detail in the footnotes to the tables.

[25] This gives a total of 20 terms, but 8 have their coefficient repeated, so the included variable is the sum of two original terms ($m+l$, $2ml+l^2$, m^2+2lm, $km+kl$). This gives a total of 16 coefficients to be estimated in the FOC equation (for 4 we only can estimate the sum of 2 and 2 coefficients because the variable is the same). We should estimate a form of the type $\ln(a+b+x\beta)$, where a

Instruments (in addition to the constant and time dummies), are a complete polynomial of order three in k_{-1}, l_{-1} and m_{-1} and the powers k, k^2 and k^3. Taking into account the autoregressive parameter, this gives two overidentifying restrictions (PF and FOC), and three when there is no quasidifferentiation ($FOCL$). In SYS we repeat this set of instruments for each equation. As the number of parameters (different from constant and dummies) only increases in one autoregressive parameter, this provides 21 additional overidentifying restrictions. In the two robustness checks, we add the extra variables considered in the FOC to the equation-specific set of instruments.

Table 2 reports the probability values of the specification test (χ^2 or Sargan test) of the estimates. We consider incidentally what happens with the specification test when we use l and m as instruments.

Single equation estimates

Tables 2 and 3 show in the first place that the FOC constitutes an alternative to the production function in order to estimate the markups. Correlation is not perfect, but quite strong. Markups computed with the level of the FOC show a 0.5 correlation with the markups computed from the production function elasticities. And the markups computed with the level of the FOC and the FOC differentiated show a 0.65 correlation. The dispersion of the values of the markups are, however, quite different. Dispersion tends to be greater with the markups computed with the production function, smaller when the FOC is used, and intermediate when the FOC is differentiated. How to interpret the differences?

A rough rule to value the components of the computed markups is to think of $\ln(R/VC) + \ln \widehat{\nu}$ as approximable by $(R - VC)/VC - (1 - \widehat{\nu})$. For a given value of

and b are constants. We use the approximation $\ln(a + b + x\beta) = \ln(a + b) + \ln(1 + x\beta/(a + b)) \simeq \ln(a + b) + x\beta/(a + b)$. This avoids a nonlinear search on 14 additional coefficients. Parameters a and b are obviously entirely identified by the production function.

the price average cost margin, the markup will be smaller the lower the computed elasticity ν (MC is larger in relation to AVC). The production function estimate produces a great dispersion of the elasticities, some of them implausibly low so that the proportion of negative markups tends to be too high. Something similar happens with the elasticities produced by the quasidifferentiation of the FOC, but not with the own FOC. Our conclusion is that the FOC contains information that can help to estimate the elasticities of scale, probably because is less sensitive to specification errors than the production function, but the quasidiffereciation exacerbates the variance rendering difficult to take advantage of this characteristic.

The values of the specification tests point out at the validity of the used instruments. The specification tests are all passed with high probability values in the estimation of the production function. When we estimate the elasticities from the level of the FOC, all tests but industry 4 are passed, although with more moderate probability values. And when we estimate from the differentiated FOC, all of them again are passed with no exception and with high probability values. To check that this is not resulting from a low power of the test we repeat the estimation from the levels of the FOC replacing the instruments k^2 and k^3 by l and m. The result, reported in column (5) of Table 2, is that the test is not passed at the 5% of significance in 8 out of the 10 industries.

Our conclusion from the single-equation estimates is that the distribution of elasticities of scale seem to be in general consistently estimable both from the production function and the FOC, using as moments the same that the context of the production function suggests (k, k_{-1}, l_{-1} and m_{-1}). The level of the FOC provides a nice estimation of the distribution of elasticities, but also depends on correlation assumptions that are more difficult to sustain and less neatly accepted by the data. The production function and differentiated FOC estimates are noisier, but rely on very likely correlation assumptions.

In general, a big advantage of estimating from only the FOC is that we do not need to make any assumption on the process of productivity. In fact is quite remarkable that, under simple assumptions of correlation of K, L and M with $\ln \mu$, we are able to estimate the elasticities of scale up to a constant without output prices. This is not possible with the production function.[26]

If a researcher basically needs to test the association of any variable with the markups, asses their evolution over time, or compare their value across groups of firms, these results suggest to run a regression of $\ln(R/VC)$ on a constant and time dummies, the corresponding variable or dummies of interest, controlling by a polynomial on K, L and M to account for the differences between AVC and MC. Running the regression in levels and quasidifferentiating, as well as computing the specification tests, seems the most advisable way to do the things. In general, no individual markups nor the markup levels are going to be estimated. However, the researcher can claim that her inference results are robust to the form of competition among firms, to the functional form of the technology, and consistent. This way to proceed doesn't need the availability of output prices. We comment later about what to do if one is suspicious about the presence of labor augmenting productivity.

Joint estimation

If estimating the elasticities of scale can be done consistently from both the production function and the first order condition, estimating jointly can produce advantages. First, the effects of L and M are now going to be subject to the requirement to explain the output at the same time that their derivatives have to explain part of the observed margins. This can limit the dispersion that can emerge because errors in variables or in specification. On the other hand, the estimation provides naturally an estimate for the constant of the elasticity of scale. In practice, Table 4 shows that the joint

[26] As remarked by Bond, Hashemi, Kaplan and Zoch (2020).

estimation of the FOC and the production function in SYS gives perfectly comparable average markups. A little surprisingly, since both the production function and the FOC are now quasidifferentiated, the estimated markups show a correlation of almost 0.9 with the estimates from the levels of the FOC. The distribution of elasticities shows a very reasonable dispersion and, as a result, the proportion of estimated negative markups is completely sensible.

The price of these gains is that we are assuming a linear dynamic specification of productivity in the production function. In fact, one may be worried that, although the probability values of the specification test are good, the test is not passed at the 5% level of significance in industries 2 and 7. This may be related to the fact that we are now checking a broader specification (validity of the moments for the two equations simultaneously) with higher efficiency. A clear way to check if there is some correlation left is to include in the specification the possibly omitted variables that can be creating this correlation. The robustness checks reported in columns (10) to (13) show that including markup shifters all the specification tests tend to be passed again with almost no change in the estimates. Dispersion of the elasticities is relatively good in the first robustness check and worse in the second (as could be expected by its nature).

Some examples from DLW estimators

We close this empirical exercise showing some characteristics of the DLW estimates, revealed by applying them to the same samples. This is done in Table 4, while the correlation among the different DLW estimates and between the DLW estimates and the consistent estimates are reported in Table 5. We estimate using the most standard instruments, contemporaneous capital and lagged variable inputs. The first observation is about the requirements of the estimators. Our consistent proposals do not need the use of input prices (only the observation of VC). Column (1) of Table 4 shows that, without including input prices in the first stage, a simple DLW estimator using the Cobb-Douglas production function produces half of the markup means too high (around 0.4) or too low (around zero). The second observation is that using a completely flexible functional form is definitely important for robustness. Columns (2) to (5) of Table 4, which include two CD and two Translog (TL) estimates, show that the DLW markup estimates are very sensitive to the functional form. Table 5 shows correlations among the DLW estimators in the range 0.5-0.6, but CD can produce a zero mean and TL negative means (additional information shows that CD tend to produce no negative markups and TL a high proportion). Notice that, somewhat strikingly, the TL estimates show very low correlations with the consistent estimates (in the range 0.2 to 0.4).

The effects of the main criticisms addressed by this note are summarized in columns (6) to (9) of Table 4. Take a variable truly positively correlated with the markup and do-not-include/include it by turn in the first stage regression of DLW. We do this with the firm-level indicator of the state of demand. Columns (6) and (7) then show the effect of the markups computed in this way when regressed on the indicator of the state of demand. The estimated markups tend to be negatively and positively correlated respectively with this variable (we go from two significant negative effects to nine significant positive effects).

Take now an effect that represents a likely surprise. Something that is probably scarcely correlated with other included variables. We chose the last observed markup for firms that shut down the following year. And we do not include the corresponding dummy in the first stage of DLW. Column (8), using the consistent estimator $FOCL$, assess the average markup of the year before death as significantly negative in seven industries (by about -11 percentage points). The DLW TL estimator only picks this effect correctly in two cases, gives two more significant measurements totally reversed (2.4 and 9.8 percentage points), and values the margin of the exitor in its last year as non-distinguishable from zero in the rest.

6. Three extensions

Heterogeneity in the efficiency of the labor input and its increase over time is a traditional and powerful candidate to explain systematically varying labor elasticities and hence the change in relative levels of all elasticities. Broadly defined, labor-augmenting productivity includes everything that may increase marginal productivity and wages, even if it is not strictly technology. Labor-augmenting productivity, under limited elasticity of substitution, is an explanation for the decreasing trend found in the labor share at both macro and micro levels in many countries and samples, and it has been claimed the origin of mismeasurements of the markups (Doraszelski and Jaumandreu 2018 and 2019; Raval, 2019; Demirer, 2020).

The framework that we have presented seems in principle largely suitable for addressing the presence of labor-augmenting productivity, but most of the work still remains to be done. For example, the nonparametric modeling of the short-run elasticity of scale, both in the production function and the FOC, can be completed with systematic trends to be empirically measured affecting the effect of observed labor. Also, some nonparametric procedure adapted from the parametric substitution solutions of the previous papers of the authors (based on the optimal ratio of variable inputs) can be tried. We think that this is an important direction for research.

We have seen that, with markups varying in an unspecified way across firms and over time, there is no available a OP/LP solution to the estimation of the elasticity of scale or elasticities of the inputs. This makes more stringent the assumptions needed to estimate markups. Here we briefly suggest two methods that, by adding another relationship to the system explored in section 5, can help to develop more general solutions to estimate consistently markups. The first one is the adaptation of the OP/LP type of procedure. The second relies on strengthening the assumptions on the stochastic properties of μ.

The first way is estimating $\nu(K, L, M)$ using a "proxy" to control for the markup. That is, use an endogenous variable functionally related to the markups by an invertible relationship (to mimic the OP/LP method that "proxies" unobserved productivity by inverting the demand for an input). Write the demand of the firm as $Q^* = D(P, A, z)$, where A is a fixed outlay and z represents all the rest of demand variables. Profits are $\pi = PD(P, A, z) - C(D(P, A, z)) - A$. Any fixed monetary outlay A that shifts demand with time limited effects is a candidate because its optimal level is going to be determined (in the static case) according to the FOC

$$P(1 - \frac{1}{\mu})\frac{\partial Q^*}{\partial A} = 1.$$

Noticing that $P\frac{\partial Q^*}{\partial A} = \frac{\partial (pQ^*)}{\partial A}$ for a given price, we can easily deduce the quasilinear relationship

$$\mu = \frac{1}{1 - (1/\frac{\partial (pQ^*)}{\partial A})}.$$

Let us write, in completely general terms, $\ln \mu = h(A, z_A)$, where z_A is the subset of demand variables that is relevant for the determination of the sensitivity of revenue to A. If we are willing to assume that μ is a "scalar unobservable" in this relationship, i.e. there are no other unobservables in z_A, the regression

$$\ln \frac{R}{VC} = -\ln \nu(K, L, M) + h(A, z_A) + \varepsilon$$

can estimate both $\nu(K, L, M)$, $\ln \mu$ and ε without imposing any restriction on ω. The simultaneous estimation of the production function imposing the constraints $\ln \nu(K, L, M) = \ln(\frac{\partial \ln F}{\partial l} + \frac{\partial \ln F}{\partial m})$ would increase efficiency, but it is not strictly needed. If the scalar unobservable assumption is difficult to sustain everything becomes more difficult. An obvious generalization of this model is adopting a dynamic framework for the effect of A, but this raises rapidly the concerns about other unobservables.

The second way to control for $\ln \mu$ is to fully develop the idea that it is driven by a first degree stochastic difference equation with shifters, $\ln \mu = h(\ln \mu_{-1}, S)$, where S are the relevant shifters. Pricing is dynamic. The firm must have a policy about how to deal with the input prices, productivity and demand shocks and redress the markup towards long-run values and this is what the stochastic difference equation reflects (Jaumandreu and Lin, 2018).

For example, in a dynamic pricing model, in which the firm chooses the markup subject to quadratic costs of adjustment (because the cost of changing prices and other), the firm finds optimal to adjust period to period one fraction of the gap between the markup adjusted for the changes induced by prices and productivity (current and expected) and the long-run targeted markup. The way that the shifters should enter depend on their stochastic specification. We have implicitly used a model of this type by assuming an $AR(1)$ and including in the shifters the variations of capital, input prices and an indicator of the changes in demand, and checking if the quasidifferentiation of $\ln \mu$ gives a reasonable model. A more structural model provides an additional relationship between the unobservables that can be used to replace them by observables.

7. Concluding remarks

Economists would like to have, for theoretical and policy discussion purposes, easy-to-compute estimates of firm-level markups that are independent on behavior and technological assumptions. Bain (1951) ratio revenue to variable cost is close to such ideal but substitutes average variable cost for marginal cost, what makes it in general suspicious of mismeaurement. The production approach to compute markups has searched out ways to solve this problem, and De Loecker and Warzynski (2012) proposed to use an empirical measure of the ratio of the production elasticity of one variable input to its share in (corrected) revenue. The application of this measurement, however, has raised a series of micro and macro evidences at odds with economic theory and contested by other evidences.

We have argued that this is due to several problems. First, DLW method draws heavily on imposing a functional form to the production function without paying attention at the observations that can give us an indication if this is sensibly done. Second, the method is circular, in the sense that to estimate consistently the elasticity of an input and the disturbance of the production function -under unspecified imperfect competition- we would need to know marginal cost (that it is what we are looking for). Third, both used functional forms and the estimation method collapse in the presence of non-neutral technological change, a kind of technological progress that researchers are discovering to play an important role in modern production.

We build alternatives starting from the production function and the first order conditions of cost minimization. We propose to take the observed ratio revenue to variable cost, and try to separate from it what is the variation of the ratio average variable cost to marginal cost. Estimating the ratio average variable to marginal cost can be robustly done with the help of a nonparametric production function and controlling for the possible correlation between the inputs and the markup to estimate. Averaging across uncorrelated errors, this provides a nice alternative to consistently measure the impact of variables or time on the markups, both as an integrated or several steps exercise.

Appendix A: Bias formula

Consider the IV estimator of β in $\overline{q} = \beta\widetilde{l} + \rho r_{-1} + \xi + \varepsilon$, where $\widetilde{l} = l - \rho l_{-1}$, and using the instrument l uncorrelated with ξ and ε but correlated with r_{-1}. $E(l\overline{q}) = \beta E(l\widetilde{l}) + \rho E(lr_{-1}) + E(l\xi) + E(l\varepsilon) = \beta E(l\widetilde{l}) + E(lr_{-1})$, and, if $E(lr_{-1}) = Cov(l, r_{-1})$ and $E(l\widetilde{l}) = Cov(l, \widetilde{l})$,

$$\beta_{IV} = \beta(1 + \frac{1}{\beta}\frac{\sigma_r}{\sigma_{\widetilde{l}}}\frac{Corr(l, r_{-1})}{Corr(l, \widetilde{l})}),$$

where $Corr(l, r_{-1}) = \frac{Cov(l, r_{-1})}{\sigma_l \sigma_r}$ and $Corr(l, \widetilde{l}) = \frac{Cov(l, \widetilde{l})}{\sigma_l \sigma_{\widetilde{l}}}$.

On the other hand, $Var(\overline{q} - \xi - \varepsilon) = \beta^2 Var(\widetilde{l}) + \rho^2 Var(r) + 2\beta Cov(\widetilde{l}, r_{-1})$. Solving for β and rearranging a little we have

$$\beta = (\sqrt{Corr(\widetilde{l}, r_{-1})^2 + \frac{R^2}{1 - R^2}} - Corr(\widetilde{l}, r_{-1}))\frac{\sigma_r}{\sigma_{\widetilde{l}}},$$

where $R^2 = \frac{Var(\overline{q} - \xi - \varepsilon) - Var(r)}{Var(\overline{q})}$ and $Corr(\widetilde{l}, r_{-1}) = \frac{Cov(\widetilde{l}, r_{-1})}{\sigma_{\widetilde{l}} \sigma_r}$. Combining both formulas give

$$\beta_{IV} = \beta \left(1 + \rho \frac{1}{\sqrt{Corr(\widetilde{l}, r_{-1})^2 + \frac{R^2}{(1 - R^2)}} - Corr(\widetilde{l}, r_{-1})} \frac{Corr(l, r_{-1})}{Corr(l, \widetilde{l})} \right).$$

References

Ackerberg, D. (2020), "Timing Assumptions and efficiency: Empirical evidence in a Production Function Context," mimeo, University of Texas Austin.

Ackerberg, D., K. Caves and G. Frazer (2015), "Structural Identification of Production Functions," *Econometrica*, 6, 2411-2451.

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

Autor, D., D. Dorn, L. Katz, C. Patterson and J. Van Reenen (2020), "The Fall of the Labor Share and the Rise of Superstar Firms," forthcoming, *Quarterly Journal of Economics*.

Bain, J. (1951), "Relation of Profit Rate to Industry Concentration: American Manufacturing, 1936-1940," *Quarterly Journal of Economics*, 65, 293-324.

Berry, S., J. Levinsohn and A. Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 841-890.

Berry, S., M. Gaynor and F.S. Morton (2019), "Do Increasing Markups Matter? Lessons from Empirical Industrial Organization", *Journal of Economic Perspectives*, 33, 44-68.

Blum, B., S. Claro, I. Horstmann and D. Rivers (2018), "The ABCs of Firm Heterogeneity: The Effects of Demand and Cost Differences on Exporting," mimeo, University of Toronto.

Blundell, R. and S. Bond (2000), "GMM Estimation with Persistent Panel Data: An Application to Production Functions," *Econometric Reviews* 19, 321-340.

Bond, S., A. Hashemi, G. Kaplan, and P. Zoch (2020), "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," mimeo, University of Chicago.

Brandt, L., J. Van Biesebroeck, L. Wang and Y. Zhang (2017), "WTO Accession and Performance of Chinese Manufacturing Firms," *American Economic Review,* 107, 2784-2820.

Brandt, L., J. Van Biesebroeck, L. Wang and Y. Zhang (2019), "WTO Accession and Performance of Chinese Manufacturing Firms: Corrigendum," *American Economic Review,* 109, 1616-1621.

De Loecker, J, and J. Eeckhout (2018), "Global Market Power," NBER Working Paper No. 24768.

De Loecker, J., J. Eeckhout and G. Unger (2020), "The Rise of Market Power and the Macroeconomic Implications," forthcoming, *Quarterly Journal of Economics.*

De Loecker, J., P. Goldberg, A. Khandelval and N. Pavcnik (2016), "Prices, Markups and Trade Reform," *Econometrica,* 84, 445-510.

De Loecker, J. and P. Scott (2016), "Estimating Market Power: Evidence from the US Brewing Industry," NBER Working Paper No. 22957.

De Loecker, J. and F. Warzynski (2012), "Markups and Firm-level Export Status," *American Economic Review*, 102, 2437-2471.

Demirer, Mert (2020), "Production function Estimation with Factor-Augmenting Technology: An Application to Markups," mimeo, MIT.

Doraszelski, U. and J. Jaumandreu (2013), "R&D and Productivity: Estimating Endogenous Productivity," *Review of Economic Studies,* 80, 1338-1383.

Doraszelski, U. and J. Jaumandreu (2018), "Measuring the Bias of Technological Change," *Journal of Political Economy,* 126, 1027-1084.

Doraszelski, U. and J. Jaumandreu (2019), "Using Cost Minimization to Estimate Markups," CEPR Discussion Paper 14114.

Eslava, M. and J. Haltiwanger (2020), "The Life-Cycle Growth of Plants: The Role of Production, Demand, Wedges," NBER Working Paper No.27184.

Hall, R.E. (1988), "The relation between price and marginal cost in US industry,"*Journal of Political Economy*, 96, 921-947.

Jaumandreu, J. and S. Lin (2018), "Pricing Under Innovation: Evidence from Manufacturing Firms," CEPR Discussion Paper 13146.

Jaumandreu, J. and H. Yin (2020), "Cost and Product Advantages: Evidence from Chinese Manufacturing Firms," CEPR Discussion Paper 11862.

Levinsohn, J and A. Petrin (2003), "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies,* 70, 317-341.

Olley, S. and A. Pakes (1996), "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, 64, 1263-1298.

Raval, D. (2019), "The Micro Elasticity of Substitution and Non-neutral Technology," RJE, 50, 147-167.

Raval, D. (2020), "Testing the Production Approach to Markup Estimation," mimeo, FTC.

Samuelson, P. (1947), *Foundations of Economic Analysis*, Harvard University Press, Cambridge.

Table 2: Consistent markup estimates.

	Single-equation estimates							Two-equation estimates					
	PF Markup (s. dev.)	S. test P-value (2 df.)	FOCL Markup (s. dev.)	S. test P-value (3 df..)	S. test P-value (3 df.)	FOC Markup (s. dev.)	S. test P-value (2 df.)	SYS Markup (s. dev.)	S. test P-value (23 df.)	CSYS Markup (s. dev.)	S. test P-value (22 df.)	SYSMC Markup (s. dev.)	S. test P-value (14 df.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
1. Metals and metal products	0.116	0.767	0.115	0.157	0.000	0.115	0.898	0.172	0.445	0.171	0.654	0.102	0.594
	(0.608)		(0.154)			(0.280)		(0.158)		(0.158)		(0.195)	
2. Non-metallic minerals	0.235	0.838	0.235	0.286	0.006	0.235	0.448	0.211	0.029	0.208	0.121	0.170	0.467
	(0.262)		(0.192)			(0.508)		(0.203)		(0.200)		(0.368)	
3. Chemical products	0.213	0.547	0.213	0.109	0.196	0.213	0.923	0.305	0.168	0.281	0.906	0.285	0.318
	(0.288)		(0.202)			(0.258)		(0.208)		(0.215)		(0.222)	
4. Agric. and ind. machinery	0.214	0.821	0.214	0.022	0.001	0.214	0.804	0.215	0.370	0.214	0.293	0.258	0.728
	(0.322)		(0.160)			(0.214)		(0.208)		(0.210)		(0.352)	
5. Electrical goods	0.312	0.608	0.312	0.323	0.000	0.312	0.753	0.270	0.182	0.259	0.420	0.301	0.382
	(0.303)		(0.194)			(0.227)		(0.197)		(0.208)		(0.219)	
6. Transport equipment	0.161	0.986	0.161	0.113	0.003	0.161	0.912	0.296	0.157	0.166	0.324	0.179	0.431
	(0.274)		(0.171)			(0.225)		(0.185)		(0.205)		(0.297)	
7. Food, drink and tobacco	0.228	0.698	0.229	0.111	0.000	0.229	0.893	0.173	0.015	0.077	0.028	0.221	0.483
	(0.336)		(0.198)			(0.608)		(0.247)		(0.352)		(0.277)	
8. Textile, leather and shoes	0.123	0.683	0.123	0.183	0.000	0.123	0.623	0.168	0.100	0.168	0.130	0.175	0.579
	(0.644)		(0.156)			(0.167)		(0.159)		(0.158)		(0.177)	
9. Timber and furniture	0.034	0.869	0.034	0.219	0.019	0.034	0.526	0.189	0.129	0.184	0.571	0.166	0.439
	(0.678)		(0.159)			(0.226)		(0.168)		(0.244)		(0.330)	
10. Paper and printing products	0.224	0.832	0.224	0.053	0.286	0.224	0.968	0.303	0.491	0.299	0.447	0.230	0.535
	(0.589)		(0.184)			(0.293)		(0.188)		(0.193)		(0.333)	

Table 3: Correlation between consistent markup estimates[a]

	PF	FOCL	FOC	SYS	CSYS
FOCL	0.481				
FOC	0.348	0.647			
SYS	0.501	0.896	0.578		
CSYS	0.515	0.833	0.542	0.952	
SYSMC	0.556	0.670	0.469	0.699	0.706

[a] Average of the coefficients of correlation in the 10 industries.

Table 4: DLW markups.

	Mean markups					Markups regressed on md (incl. time dummies)		Markups regressed on dummy of last year of exitor (incl. t.d.)	
	DLW CD no input prices (s. dev)	DLW CD non-incl. md (s. dev.)	DLW CD incl. md (s. dev.)	DLW TL non-incl. md (s. dev.)	DLW TL incl. md (s. dev.)	DLW CD non-incl. md (s. e.)	DLW CD incl. md (s. e.)	FOCL estimation (s. e.)	DLW TL incl. md (s. e.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Metals and metal products	0.175	0.297	0.285	-0.278	0.223	-0.017*	0.036*	-0.134*	0.017
	(0.164)	(0.091)	(0.095)	(0.707)	(0.124)	(0.006)	(0.007)	(0.024)	(0.015)
2. Non-metallic minerals	0.206	0.242	0.244	0.182	0.075	-0.004	0.067*	-0.101*	0.006
	(0.170)	(0.118)	(0.125)	(0.179)	(0.313)	(0.012)	(0.015)	(0.039)	(0.038)
3. Chemical products	0.406	0.280	0.295	0.243	0.253	0.014	0.060*	-0.041	0.043
	(0.181)	(0.166)	(0.169)	(0.272)	(0.284)	(0.017)	(0.018)	(0.026)	(0.034)
4. Agric. and ind. machinery	0.408	0.252	0.261	0.238	-0.187	-0.007	0.024*	-0.119*	0.029
	(0.140)	(0.108)	(0.109)	(0.352)	(0.595)	(0.011)	(0.011)	(0.048)	(0.108)
5. Electrical goods	-0.042	0.169	0.164	0.656	0.461	-0.005	0.035*	-0.037	-0.067
	(0.181)	(0.156)	(0.157)	(0.216)	(0.274)	(0.017)	(0.018)	(0.033)	(0.039)
6. Transport equipment	0.143	0.258	0.252	0.162	0.140	0.012	0.061*	-0.101*	-0.107*
	(0.161)	(0.129)	(0.133)	(0.181)	(0.175)	(0.014)	(0.015)	(0.033)	(0.035)
7. Food, drink and tobacco	0.152	0.131	0.218	0.187	0.291	-0.066*	-0.040*	-0.143*	0.098*
	(0.200)	(0.170)	(0.172)	(0.255)	(0.368)	(0.014)	(0.014)	(0.036)	(0.051)
8. Textile, leather and shoes	0.025	0.009	0.022	0.162	0.150	-0.002	0.049*	-0.070*	0.024*
	(0.154)	(0.091)	(0.096)	(0.208)	(0.132)	(0.007)	(0.008)	(0.015)	(0.012)
9. Timber and furniture	0.249	0.102	0.103	0.146	0.178	-0.012	0.045*	-0.094*	-0.145*
	(0.132)	(0.102)	(0.107)	(0.294)	(0.275)	(0.010)	(0.011)	(0.028)	(0.040)
10. Paper and printing products	0.380	0.282	0.242	-0.026	0.036	-0.032	0.024	-0.014	0.011
	(0.185)	(0.150)	(0.152)	(0.401)	(0.264)	(0.017)	(0.017)	(0.025)	(0.035)

Table 5: Correlation among DLW firm-level markup estimates and with consistent estimates[a]

	DLW CD no	DLW CD md	DLW TL no	PF	FOCL	SYS
DLW CD no				0.401	0.519	0.593
DLW CD md	0.970			0.409	0.546	0.614
DLW TL no	0.547	0.532		0.216	0.257	0.271
DLE TL md	0.607	0.621	0.659	0.310	0.309	0.398

[a] Average of the coefficients of correlation in the 10 industries.

Table A2.1: Estimated autoregressive parameters.

	Single-equation estimates		Two-equation estimates					
	PF	FOC	SYS		C. SYS		SYS MC	
	ρ_ω (std. e.)	ρ_μ (std. e.)	ρ_ω (std. e.)	ρ_μ (std. e.)	ρ_ω (std. e.)	ρ_μ (std. e.)	ρ_ω (std. e.)	ρ_μ (std. e.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1. Metals and metal products	0.820	0.951	0.563	0.651	0.589	0.668	0.545	0.524
	(0.086)	(0.053)	(0.122)	(0.058)	(0.122)	(0.057)	(0.131)	(0.121)
2. Non-metallic minerals	0.551	0.957	0.797	0.512	0.797	0.483	1.017	0.760
	(0.071)	(0.030)	(0.056)	(0.075)	(0.051)	(0.075)	(0.014)	(0.052)
3. Chemical products	0.781	1.002	0.780	0.779	0.793	0.787	0.762	0.762
	(0.656)	(0.065)	(0.054)	(0.045)	(0.050)	(0.045)	(0.138)	(0.061)
4. Agric. and ind. machinery	0787	0.613	0.709	0.718	0.720	0.718	0.777	0.816
	(0.089)	(0.173)	(0.065)	(0.067)	(0.062)	(0.070)	(0.065)	(0.035)
5. Electrical goods	0.670	1.119	0.524	0.754	0.567	0.765	1.006	0.599
	(0.163)	(0.061)	(0.190)	(0.052)	(0.196)	(0.049)	(0.007)	(0.096)
6. Transport equipment	0.262	0.801	0.963	0.936	0.858	0.867	0.850	0.857
	(0.142)	(0.064)	(0.034)	(0.035)	(0.043)	(0.045)	(0.047)	(0.034)
7. Food, drink and tobacco	0.779	0.964	1.008	0.855	1.000	0.885	0.325	0.807
	(0.144)	(0.012)	(0.003)	(0.028)	(0.003)	(0.023)	(0.174)	(0.051)
8. Textile, leather and shoes	0.807	0.625	0.659	0.611	0.646	0.608	0.865	0.560
	(0.0.30)	(0.060)	(0.069)	(0.067)	(0.070)	(0.067)	(0.089)	(0.072)
9. Timber and furniture	0.784	0.553	0.476	0.594	0.663	0.645	0.638	0.697
	(0.040)	(0.106)	(0.059)	(0.063)	(0.051)	(0.054)	(0.066)	(0.043)
10. Paper and printing products	0.788	0.788	0.636	0.762	0.767	0.764	0.895	0.762
	(0.023)	(0.028)	(0.140)	(0.044)	(0.061)	(0.040)	(0.025)	(0.045)

Table A2.2: Average estimate of short-run elasticity of scale.

	Single equation estimates						Two-equation estimates					
	PF		FOCL		FOC		SYS		C. SYS		SYS MC	
	ν (s. dev.)	Prop. neg.	ν (s. dev.)	Prop. neg.	ν (s. dev.)	Prop. neg.	ν (s. dev.)	Prop. neg.	ν (s. dev.)	Prop. neg.	ν (s. dev.)	Prop. neg.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1. Metals and metal products	1.036 (0.462)	0.352	0.911 (0.049)	0.158	0.937 (0.242)	0.372	0.962 (0.028)	0.087	0.962 (0.026)	0.081	0.904 (0.116)	0.296
2. Non-metallic minerals	1.024 (0.202)	0.168	1.009 (0.118)	0.086	1.114 (0.504)	0.276	0.984 (0.083)	0.126	0.974 (0.080)	0.132	0.985 (0.289)	0.292
3. Chemical products	0.955 (0.200)	0.175	0.933 (0.076)	0.101	0.943 (0.163)	0.191	1.020 (0.046)	0.018	0.999 (0.085)	0.049	1.001 (0.058)	0.054
4. Agric. and ind. Machinery	0.993 (0.226)	0.233	0.966 (0.052)	0.059	0.974 (0.136)	0.157	0.973 (0.113)	0.127	0.972 (0.116)	0.128	1.048 (0.275)	0.212
5. Electrical goods	1.056 (0.221)	0.153	1.036 (0.078)	0.028	1.046 (0.165)	0.067	0.0993 (0.063)	0.051	0.986 (0.098)	0.081	1.026 (0.086)	0.057
6. Transport equipment	0.975 (0.163)	0.191	0.960 (0.074)	0.111	0.967 (0.145)	0.209	1.009 (0.086)	0.041	0.965 (0.082)	0.179	0.993 (0.178)	0.222
7. Food, drink and tobacco	0.995 (0.225)	0.234	0.978 (0.160)	0.074	1.131 (0.592)	0.387	0.930 (0.152)	0.232	0.865 (0.230)	0.432	0.971 (0.123)	0.197
8. Textile, leather and shoes	1.100 (0.478)	0.287	0.948 (0.053)	0.150	0.948 (0.063)	0.194	0.991 (0.034)	0.088	0.991 (0.031)	0.086	1.001 (0.087)	0.128
9. Timber and furniture	1.019 (0.491)	0.367	0.864 (0.055)	0.412	0.875 (0.157)	0.485	1.009 (0.045)	0.079	1.016 (0.168)	0.221	1.017 (0.225)	0.235
10. Paper and printing products	1.054 (0.381)	0.187	0.946 (0.079)	0.057	0.972 (0.852)	0.206	1.021 (0.033)	0.018	1.017 (0.041)	0.020	0.994 (0.316)	0.245

Table A2.3: Regressing the estimated markups on the state of demand.

	Single equation estimates			Two-equation estimates		
	PF Coeff. (s. e.)	FOCL Coeff. (s. e.)	FOC Coeff. (s. e.)	SYS Coeff. (s. e.)	C. SYS Coeff. (s. e.)	SYS MC Coeff. (s. e.)
	(1)	(2)	(3)	(4)	(5)	(6)
1. Metals and metal products	-0.072	0.052	0.050	0.042	0.042	0.044
	(0.047)	(0.009)	(0.020)	(0.009)	(0.009)	(0.013)
2. Non-metallic minerals	0.057	0.071	0.261	0.045	0.043	-0.022
	(0.029)	(0.019)	(0.054)	(0.022)	(0.021)	(0.041)
3. Chemical products	0.049	0.070	0.079	0.056	0.045	0.052
	(0.028)	(0.021)	(0.026)	(0.022)	(0.022)	(0.023)
4. Agric. and ind. Machinery	-0.020	0.018	-0.012	0.017	0.015	-0.041
	(0.030)	(0.016)	(0.019)	(0.021)	(0.021)	(0.036)
5. Electrical goods	0.050	0.045	0.069	0.041	0.031	0.039
	(0.028)	(0.018)	(0.021)	(0.019)	(0.019)	(0.020)
6. Transport equipment	0.059	0.060	0.064	0.074	0.047	0.030
	(0.031)	(0.017)	(0.022)	(0.019)	(0.022)	(0.032)
7. Food, drink and tobacco	-0.116	-0.013	-0.114	0.040	0.056	-0.096
	(0.026)	(0.015)	(0.054)	(0.021)	(0.032)	(0.021)
8. Textile, leather and shoes	-0.098	0.059	0.033	0.041	0.045	0.042
	(0.059)	(0.011)	(0.012)	(0.012)	(0.011)	(0.013)
9. Timber and furniture	-0.267	0.047	0.011	0.025	-0.062	-0.061
	(0.070)	(0.014)	(0.022)	(0.015)	(0.022)	(0.030)
10. Paper and printing products	0.012	0.040	0.053	0.034	0.027	0.077
	(0.064)	(0.016)	(0.032)	(0.017)	(0.017)	(0.039)