Reexamining the De Loecker & Warzynski (2012) method for estimating markups*

Ulrich Doraszelski
University of Pennsylvania†

Jordi Jaumandreu
Boston University‡

October 30, 2023

Abstract

De Loecker & Warzynski (2012) obtain the markup from the firm's cost minimization problem by substituting in estimates of the output elasticity of a variable input and the disturbance that separates actual from planned output. We show, however, that consistently estimating the output elasticity and the disturbance using the procedure developed by Olley & Pakes (1996) and Levinsohn & Petrin (2003) generally requires observing and controlling for the markup. We analyze the resulting biases and discuss alternative approaches to estimation.

*An earlier draft of this paper was circulated in 2020 under the title "The inconsistency of De Loecker & Warzynski's (2012) method to estimate markups and a robust alternative". We are grateful to Ivan Fernandez-Val, Jiaoyang Huang, and Lixiong Li for useful discussions and to Mike Abito, Mert Demirer, Rigoberto Lopez, and Katja Seim for comments. Hyun Soo Suh and Guilherme Neves Silveira provided excellent research assistance. Doraszelski gratefully acknowledges financial support from the Dean's Research Fund and the AI and Analytics for Business Program at the Wharton School.

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

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

1 Introduction

Following a wave of acquisitions in an industry, a policymaker asks an economist if, and by how much, market power has increased. To answer this question, the economist has detailed production data for a sample of firms in the industry at her disposal, including output and input quantities and prices. The economist uses the De Loecker & Warzynski (2012) (henceforth DLW) method to estimate the firm-level markup and then regresses it on a step dummy that is one for the acquiring firms in the "after" period and zero otherwise. If the acquisitions have in truth raised the markup of these firms, this regression is likely to tell the opposite. In this paper, we explain what has gone wrong in the exercise of the economist and how to address the problem.

To provide policy advice and answer a variety of empirical and theoretical questions, economists would like to have an easy-to-compute way to estimate the firm-level markup that does not require modelling demand and making assumptions about firm conduct. Bain's (1951) ratio of revenue to variable cost comes close to this ideal but relies on equating inherently unobservable marginal cost with average variable cost.

The production approach to estimating the markup has searched for ways out of this impasse. DLW note that very generally a firm minimizes its cost irrespective of the specifics of demand and firm conduct. They therefore obtain the markup from the FOC for cost minimization by substituting in estimates of the output elasticity of a variable input and the disturbance that separates the firm's actual output as recorded in the data from the output that the firm planned on when it made its input decisions. To obtain the output elasticity and the disturbance separating actual from planned output, DLW use the procedure developed by Olley & Pakes (1996) and Levinsohn & Petrin (2003), implemented as suggested by Ackerberg, Caves & Frazer (2015) (henceforth OP, LP, and ACF), to estimate the production function.

The DLW method has been widely applied to study the distribution of the markup across firms and its evolution over time (see, e.g., De Loecker, Goldberg, Khandelwal & Pavcnik 2016, Brandt, Van Biesebroeck, Wang & Zhang 2017, Brandt, Van Biesebroeck, Wang & Zhang 2019, De Loecker & Scott 2016, De Loecker, Eeckhout & Unger 2018, De Loecker & Eeckhout 2018, Autor, Dorn, Katz, Patterson & Van Reenen 2020).[1] In this paper,

[1]See also Berry, Gaynor & Scott-Morton (2020) for a recent panorama of the industrial organization literature on markups.

we first characterize the circumstances under which the DLW method consistently estimates markups. We then show that outside these circumstances the DLW method produces inconsistent estimates of the output elasticity and the disturbance and therefore biased markups.[2] In particular, the DLW method is not robust to any differences in demand across firms or time unless they are observed by the econometrician in their entirety. Similarly, the DLW method is not robust to any unobserved changes in firm conduct.

This poses an especially thorny issue because, besides its conduct, the demand a firm faces in the output market is a fundamental determinant of the markup that the firm charges. At the same time, the large literatures on demand estimation and productivity analysis make clear that controlling for differences in demand by observables is a daunting task. Papers such as Berry, Levinsohn & Pakes (1995) and Foster, Haltiwanger & Syverson (2008) notably highlight the considerable heterogeneity in demand that remains even after controlling for detailed product attributes or honing in on (nearly) homogenous products. The issue is compounded by the fact that, in imperfectly competitive industries, the demand a firm faces depends on its rivals, which are partially or completely unobserved in typical production data.

The intuitive reason for the inconsistency of the DLW method is as follows. The OP/LP procedure solves the endogeneity problem in production function estimation by inverting a decision of the firm that the econometrician observes, such as the firm's demand for a variable input, to recover the firm's productivity that the econometrician does not observe. This inversion presumes that two firms that have the same productivity have the same input demand. If there is heterogeneity in demand in addition to heterogeneity in productivity, then this is not the case: two firms that have the same productivity but charge different markups because they face different demands in the output market generally have different input demands. It is therefore no longer possible to express unobserved productivity in terms of observables. Put differently, to use the OP/LP procedure to estimate the production function and obtain the markup, the DLW method would have to observe and control for the markup. In this way, the DLW method is

[2] We first noted the inconsistency of the DLW method in Doraszelski & Jaumandreu (2019), where we emphasize the implications of biased technological change for markup estimation. Bond, Hashemi, Kaplan & Zoch (2021) reiterate our point, although their focus is on the difficulties for markup estimation that arise if the econometrician observes revenue rather than the quantity of output.

circular.

The observation that the proxy variable methods developed by OP and LP cannot accommodate unobserved demand heterogeneity has been made before. Foster et al. (2008) put it as follows:

> ... idiosyncratic demand shocks make the proxies functions of both technology and demand shocks, thereby inducing a possible omitted variable bias. Put simply, proxy methods require a one-to-one mapping between plant-level productivity and the observables used to proxy for productivity. This mapping breaks down if other unobservable plant-level factors besides productivity drive changes in the observable proxy. (p. 403)

At first glance, however, this observation appears irrelevant under the cost minimization assumption of DLW. The purpose of relying on cost minimization instead of profit maximization is precisely to insulate the estimated markup from the specifics of demand and firm conduct. After all, a firm minimizes its cost in most circumstances. De Loecker et al. (2016) accordingly state that (their extension of) the DLW method (to multi-product firms) "does not require assumptions on the market structure or demand curves faced by firms" (p. 445, see also p. 497).

This is an overstatement. We show that in the cost minimization problem the firm's planned output summarizes the demand the firm faces and its conduct. Because planned output remains unobserved by the econometrician, the firm's cost-minimizing decisions again cannot be inverted to express unobserved productivity in terms of observables. The DLW method therefore either has to rule out any differences in demand and conduct across firms and time or assume that they can be fully controlled for by observables.

At a minimum, the conditions required by the DLW method to consistently estimate markups must be justified from a detailed understanding of the market structure and the demands firms face in the industry under study. This negates the purported advantage of relying on cost minimization in the production approach to estimating the markup. Whether one can fully control for any differences in demand across firms and time by observables also remains questionable in light of the large literatures on demand estimation and productivity analysis (Berry et al. 1995, Foster et al. 2008), and fully controlling for any differences in conduct may be equally challenging.

We therefore first characterize the bias in the estimates produced by the DLW method that results if there are differences in demand or conduct across firms or time that cannot be fully controlled for by observables. We show that the bias permeates the level of the estimated markup and its correlation with variables of interest such as a firm's export status or measures of trade liberalization. To show that both the level and the correlation component of the bias can be severe, we develop closed-form expressions for the bias in specific settings.

Using data from the Spanish manufacturing sector, we then test for the effects of unobserved demand heterogeneity and illustrate their consequences for the estimated markup. Our empirical application indicates that unobserved demand heterogeneity is important. The resulting bias is most pronounced in the correlation of the estimated markup with variables of interest.

We then return to the conditions required by the DLW method to consistently estimate markups. We provide a way to formally assess whether the endeavor of using observables to control for differences in demand across firms and time is successful. As may be expected, success or failure hinges on the specification of demand and assumptions on firm conduct. This reinforces our point that the DLW method does not free the researcher from having to think carefully about demand and firm conduct.

We close the paper by continuing our empirical application to illustrate the dynamic panel approach to estimation. Because the dynamic panel method avoids the inversion at the heart of the OP/LP procedure, it is robust to unobserved demand heterogeneity and changes in firm conduct.

In sum, our paper makes three main contributions. First, we highlight a not duly appreciated—and sometimes completely overlooked—assumption of the DLW method. OP and LP rule out unobserved demand heterogeneity and changes in firm conduct by assumption.[3] Indeed, this has been codified as the scalar unobservable assumption (Ackerberg, Benkard, Berry & Pakes 2007) in the subsequent literature. DLW take the OP/LP procedure to a different context that allows for heterogeneity in demand and the markup as well as changes in firm conduct without acknowledging the implication, namely that to use an OP/LP procedure to estimate the production function and

[3]LP assume a perfectly competitive industry where firms act as price takers and thus face the same horizontal demand curve (see p. 322 and Appendix A). OP rule out unobserved demand heterogeneity by assuming that any profitability differences across firms are due to differences in their capital stocks and productivities (see p. 1273). Limiting the state variables in the firm's investment policy to its own capital stock and productivity implicitly abstracts from competition between firms (see also Lemma 3 and Theorem 1 in Pakes (1994)).

obtain the markup, one would have to observe and control for the markup. Second, we contribute by developing the consequences of unobserved demand heterogeneity and changes in firm conduct and characterizing the bias in the estimates produced by the DLW method. The two-step nature of the OP/LP procedure complicates the analysis and, to our knowledge, our paper is the first to show when and how a violation of the scalar unobservable assumption in the first step causes a bias in the second step. Third, we point to the dynamic panel method and other alternative approaches to estimation that are robust to differences in demand or conduct across firms or time even if they cannot be fully controlled for by observables.

The remainder of this paper is organized as follows. In Section 2, we recall the setup and the DLW method for estimating the markup. In Section 3, we argue that it is generally not possible to express unobserved productivity in terms of observables. In Section 4, we characterize the bias in the estimates if the economist nevertheless proceeds along the lines of DLW. In Section 5, we provide an empirical application to test for the effects of unobserved demand heterogeneity. In Section 6, we provide a way to formally assess whether the endeavor of using observables to control for differences in demand across firms and time is successful. In Section 7, we continue our empirical application to illustrate the dynamic panel approach and point to other alternative approaches to estimation. In Section 8, we conclude and flag issues besides unobserved demand heterogeneity and changes in firm conduct that the production function approach to estimating the markup has to confront.

2 DLW method

Firm j produces output Q_{jt} in period t with a predetermined amount of capital K_{jt} and freely variable amounts of labor L_{jt} and materials M_{jt}.[4] The production function is

$$Q_{jt} = Q^*_{jt} \exp(\varepsilon_{jt}), \quad Q^*_{jt} = F(K_{jt}, L_{jt}, M_{jt}) \exp(\omega_{jt}), \tag{1}$$

[4]Applications of OP/LP and DLW differ in the identity of the variable input: DLW alternatively assume labor or materials to be freely variable (p. 2457), De Loecker et al. (2016) assume materials to be freely variable (p. 471), and LP assume both labor and materials to be freely variable (p. 322 and p. 339). We adopt the latter assumption merely for concreteness.

where ω_{jt} is Hicks-neutral productivity that the firm observes before it decides on variable inputs in period t but that remains unobserved by the econometrician. As usual in the literature following OP and LP, ω_{jt} is governed by a first-order Markov process with the law of motion $\omega_{jt} = E(\omega_{jt}|\omega_{jt-1}) + \xi_{jt} = g(\omega_{jt-1}) + \xi_{jt}$, where $g(\omega_{jt-1})$ is expected productivity and ξ_{jt} is the productivity innovation. The disturbance ε_{jt} accounts for the difference between the firm's actual output Q_{jt} as recorded in the data and the output Q^*_{jt} that the firm planned on when it made its input decisions. While we think of ε_{jt} as measurement error for simplicity, it can alternatively be interpreted as an unanticipated shock to output (OP, pp. 1273–1274).[5,6] In contrast to ω_{jt}, ε_{jt} is mean independent of—and therefore uncorrelated with—the inputs. Because the econometrician does not observe ε_{jt}, planned output Q^*_{jt} also remains unobserved by the econometrician.

DLW assume cost minimization in an attempt to avoid specifying demand and firm conduct (pp. 2437–2438 and p. 2443). The firm minimizes variable cost $VC_{jt} = P_{Ljt}L_{jt} + P_{Mjt}M_{jt}$, where P_{Ljt} and P_{Mjt} are the prices of labor and materials, subject to achieving its planned output Q^*_{jt}. The FOC for variable input $X_{jt} \in \{L_{jt}, M_{jt}\}$ is

$$P_{Xjt} = MC(K_{jt}, P_{Ljt}, P_{Mjt}, Q^*_{jt}, \omega_{jt}) \frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}} \exp(\omega_{jt}), \quad (2)$$

where the envelope theorem serves to replace the Lagrange multiplier by short-run marginal cost $MC(\cdot)$.

The markup is defined as $\mu_{jt} = \frac{P_{jt}}{MC(\cdot)}$, where P_{jt} is the price of output. Rewriting the FOC in equation (2) using the production function in equation (1) yields

$$\mu_{jt} = \frac{P_{jt}Q_{jt}}{P_{Xjt}X_{jt}} \frac{\frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}} X_{jt}}{F(K_{jt}, L_{jt}, M_{jt}) \exp(\varepsilon_{jt})} = \frac{\beta_X(K_{jt}, L_{jt}, M_{jt})}{S^R_{Xjt}} \exp(-\varepsilon_{jt}), \quad (3)$$

where $\beta_X(\cdot) = \frac{\partial F(\cdot)}{\partial X_{jt}} \frac{X_{jt}}{F(\cdot)}$ is the output elasticity of variable input X_{jt} and $S^R_{Xjt} = \frac{P_{Xjt}X_{jt}}{P_{jt}Q_{jt}}$ is the expenditure share of the input. Note that S^R_{Xjt} is

[5] There are few attempts to account for measurement error in inputs within the proxy variable paradigm (Kim, Petrin & Song 2016, Hu, Huang & Sasaki 2020, Collard-Wexler & De Loecker 2020).

[6] Mundlak & Hoch (1965) refer to ω_{jt} and ε_{jt} as the transmitted, respectively, untransmitted component of productivity. Examples of the untransmitted component may include machine breakdowns, labor actions, supply chain disruptions, and power outages that are not anticipated by the firm.

observed because it is based on actual output Q_{jt} rather than planned output Q^*_{jt}. DLW therefore obtain the markup μ^{DLW}_{jt} of firm j in period t by substituting estimates of $\beta_X(\cdot)$ and ε_{jt} into equation (3).

To estimate the output elasticity $\beta_X(\cdot)$ and the disturbance ε_{jt}, DLW use the procedure developed by OP and LP, implemented as suggested by ACF (p. 2442 and pp. 2444–2449).[7] After specifying the OP/LP procedure, in Section 3 we show that the inversion at the heart of this procedure fails because planned output Q^*_{jt} is unobserved by the econometrician. We also show that in the cost minimization problem the firm's planned output Q^*_{jt} summarizes the demand the firm faces and its conduct. In Section 4, we show that, as a consequence, the OP/LP procedure yields inconsistent estimates $\widehat{\beta}_X(\cdot) = \beta_X(\cdot)(1 + bias_{jt})$ and $\widehat{\varepsilon}_{jt} = \zeta_{jt} + \varepsilon_{jt}$ in the presence of unobserved demand heterogeneity. The markup $\mu^{DLW}_{jt} = \mu_{jt}(1 + bias_{jt}) \exp(-\zeta_{jt})$ estimated by the DLW method is therefore biased.

3 Inverting for unobserved productivity

The OP/LP procedure starts with a function $\omega_{jt} = h(z_{jt})$ that expresses unobserved Hicks-neutral productivity ω_{jt} by a vector of observables z_{jt}. OP use the demand for investment to invert for ω_{jt} and LP the demand for a variable input. z_{jt} correspondingly collects input quantities and prices and all other arguments of the demand that is inverted for ω_{jt}.[8] Substituting the function $\omega_{jt} = h(z_{jt})$ into equation (1) and taking logs yields

$$q_{jt} = q^*_{jt} + \varepsilon_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + h(z_{jt}) + \varepsilon_{jt} = \phi(z_{jt}) + \varepsilon_{jt}, \qquad (4)$$

where we use lowercase letters to denote logs and $\phi(\cdot)$ is an unknown function that must be estimated nonparametrically. Assuming that ε_{jt} is mean

[7] While DLW and its subsequent applications sometimes also use other estimators, the OP/LP procedure plays a central role. DLW report eight estimates based on the OP/LP procedure in their Table 2, compared to two estimates based on Hall (1986) and Klette (1999). While all results reported in De Loecker et al. (2016) are based on the OP/LP procedure, De Loecker, Eeckhout & Unger (2020) also report results based on cost shares as a robustness check. Of course, these other estimators have their own drawbacks. Hall (1986) and Klette (1999) assume that the markup is constant across firms and time and the cost share "approach relies on each input of production to be variable and for production to occur under constant returns to scale" (De Loecker et al. 2018, p. 13).

[8] Following LP, DLW rely on the demand for materials and write $\omega_{it} = h_t(m_{it}, k_{it}, z_{it})$, where i indexes firms and t periods (p. 2446). To economize on the notation, we subsume their m_{it}, k_{it}, and z_{it} into our z_{jt}.

independent of z_{jt} and carrying out the regression in equation (4) yields estimates of planned output $q^*_{jt} = \phi(z_{jt})$ and the disturbance $\varepsilon_{jt} = q_{jt} - q^*_{jt}$ that separates actual output q_{jt} from planned output q^*_{jt}.[9] This is the first step of ACF.

In a second step, the estimate of $\phi(\cdot)$ and the Markovian assumption on Hicks-neutral productivity ω_{jt} serve to estimate the production function and the implied output elasticity $\beta_X(\cdot) = \frac{\partial \ln F(\cdot)}{\partial x_{jt}}$ by carrying out the regression

$$q_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt}) + g\left(\widehat{\phi}(z_{jt-1}) - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})\right) + \xi_{jt} + \varepsilon_{jt}, \tag{5}$$

where the conditional expectation function $g(\cdot)$ is estimated nonparametrically. Note that any variable input that the firm decides on after it observes ω_{jt} is correlated with the productivity innovation ξ_{jt} and must be instrumented for.

In the arguments z_{jt} of the function $\omega_{jt} = h(z_{jt})$, DLW include any "additional variables potentially affecting optimal input demand choice" and advise that "[t]he exact variables to be included ... depend on the application but will definitely capture variables leading to differences in optimal input demand across firms such as input prices" (p. 2446). As we show below, however, another variable affecting optimal input demand is planned output Q^*_{jt}. Because planned output Q^*_{jt} is unobserved by the econometrician, the inversion at the heart of the OP/LP procedure fails. Put differently, to use this procedure to estimate the disturbance ε_{jt} or, equivalently, planned output Q^*_{jt} in the first step and the output elasticity $\beta_X(\cdot)$ in the second step, DLW would have to observe and control for planned output Q^*_{jt}.

Inverting a variable input. LP use the demand for a variable input to invert for ω_{jt}. The solution to the cost minimization problem in Section 2 is the variable cost function $VC\left(K_{jt}, P_{Ljt}, P_{Mjt}, \frac{Q^*_{jt}}{\exp(\omega_{jt})}\right)$. From Shephard's lemma, the demand for variable input X_{jt} is thus

$$X_{jt} = \frac{\partial VC\left(K_{jt}, P_{Ljt}, P_{Mjt}, \frac{Q^*_{jt}}{\exp(\omega_{jt})}\right)}{\partial P_{Xjt}}. \tag{6}$$

[9] The assumption that ε_{jt} is uncorrelated with z_{jt} slightly strengthens the earlier assumption that ε_{jt} is mean independent of the inputs.

While this expression can be inverted for $\frac{Q^*_{jt}}{\exp(\omega_{jt})}$, with planned output Q^*_{jt} being unobserved it cannot be inverted for ω_{jt}; hence, it is not possible to replace unobserved Hicks-neutral productivity ω_{jt} by observables z_{jt}. Combining the demands for two or more variable inputs does not resolve the problem.

Another way to see the problem is to go back to the FOC in equation (2). Without controlling for marginal cost $MC(K_{jt}, P_{Ljt}, P_{Mjt}, Q^*_{jt}, \omega_{jt})$, the FOC cannot be used to express unobserved Hicks-neutral productivity ω_{jt} in terms of observables z_{jt}. However, marginal cost is inherently unobservable and the assumption of cost minimization does not suffice to infer it. To see this, combine the FOCs for labor and materials and the production function in equation (1) to express marginal cost as

$$MC(K_{jt}, P_{Ljt}, P_{Mjt}, Q^*_{jt}, \omega_{jt}) = \frac{VC_{jt}}{Q^*_{jt}} \frac{1}{\nu(K_{jt}, L_{jt}, M_{jt})}, \tag{7}$$

where $VC_{jt} = P_{Ljt}L_{jt} + P_{Mjt}M_{jt}$ is variable cost and $\nu(\cdot) = \beta_L(\cdot) + \beta_M(\cdot)$ is the short-run elasticity of scale. Even though variable cost is observable, marginal cost cannot be inferred because the econometrician does not observe planned output Q^*_{jt}.[10] Planned output being unobservable is equivalent to marginal cost being unobservable.

Inverting investment. OP use the demand for investment to invert for ω_{jt}. OP derive the demand for investment from a dynamic profit maximization problem (pp. 1270–1273). This requires OP to take a stand on demand in the output market and firm conduct, which is what DLW intend to avoid. One may alternatively start from a dynamic cost minimization problem, where the firm chooses capital, labor, and materials, possibly subject to adjustment costs, to achieve a sequence of planned outputs Q^*_{jt} (see, e.g., Doraszelski & Jaumandreu 2019). In this case, the demand for investment is a function of $\frac{Q^*_{jt}}{\exp(\omega_{jt})}$ and an analogous problem to the one just reviewed arises.

Controlling for planned output. As we have shown above, as long as planned output Q^*_{jt} is unobserved by the econometrician, it is not possible

[10] The most one can infer is $MC(\cdot)\exp(-\varepsilon_{jt})$ by replacing planned output Q^*_{jt} with actual output $Q_{jt} = Q^*_{jt}\exp(\varepsilon_{jt})$ in equation (7).

to replace unobserved Hicks-neutral productivity ω_{jt} by observables z_{jt}. The obvious way around this problem is to assume that planned output Q^*_{jt} can itself be controlled for by a subset of the observables z_{jt}, i.e., that there exists a function $Q^*_{jt} = D(z^D_{jt})$ mapping observables $z^D_{jt} \subseteq z_{jt}$ into planned output Q^*_{jt}. This is the intuition behind DLW's broad interpretation of z_{jt}. De Loecker et al. (2016) include variables such as location, product dummies, export status, input and output tariffs, market share, and the price of output in z_{jt} (p. 466). Output tariffs, for example, clearly play no role in the cost minimization problem; the only reason to include them in z_{jt} is as an attempt to control for Q^*_{jt}.

The large literatures on demand estimation and productivity analysis cast doubt on any attempt to control for planned output Q^*_{jt} by observables z^D_{jt}. Berry et al. (1995) stress the importance of the unobserved characteristic that remains even after including detailed product attributes in the specification of demand. Foster et al. (2008) similarly highlight the considerable heterogeneity in demand that remains even after honing in on (nearly) homogenous products. Hence, the demand the firm faces is $Q^*_{jt} = D(z^D_{jt}, \delta_{jt})$, where the demand shock δ_{jt} captures unobserved demand heterogeneity in the sense of any differences in demand across firms or time that remain after controlling for observables z^D_{jt}.

Note that under imperfect competition δ_{jt} includes not only the unobserved product characteristic of the firm under consideration but also those of its rivals. Moreover, the demand the firm faces depends on its rivals' prices under Bertrand competition and on its rivals' (planned) quantities under Cournot competition.[11] To the extent that these variables are partially or completely unobserved, as they are in production data that covers a sample of firms, they become part of δ_{jt}.[12]

In addition to demand, planned output $Q^*_{jt} = D(z^D_{jt}, \delta_{jt})$ depends on the conduct of the firms in the industry. Changes in firm conduct, e.g., following

[11]Instead of thinking of $Q^*_{jt} = D(z^D_{jt}, \delta_{jt})$ as one of the equations in the demand system for the industry under study, we can think of $D(\cdot)$ as the firm's residual demand in the sense of Baker & Bresnahan (1985). In this case, $D(\cdot)$ encapsulates how the industry equilibrium changes as the focal firm changes its price or quantity. While this obviates accounting for rivals' prices or quantities, $D(\cdot)$ instead depends on assumptions about firm conduct and on rivals' marginal costs and thus their unobserved productivities.

[12]The common practice of letting the function $\omega_{jt} = h(z_{jt})$ vary by period may partly absorb time-series variation but of course not cross-sectional variation due to unobserved differences in demand across firms.

a wave of acquisitions as in the opening paragraph of Section 1, become part of δ_{jt} to the extent that they are unobserved and remain after controlling for observables z^D_{jt}.

In sum, in the cost minimization problem the firm's planned output $Q^*_{jt} = D(z^D_{jt}, \delta_{jt})$ summarizes the demand the firm faces and its conduct.[13] There is little reason to believe that $\delta_{jt} = 0$ as required by DLW. At the very least, assuming $\delta_{jt} = 0$ requires a careful justification starting from the specification of demand and assumptions on firm conduct, thus negating the purported advantage of the production approach and relying on cost minimization to estimate markups over the demand approach. We come back to this point in Section 6, where we provide a way to formally assess whether the endeavor of controlling for planned output Q^*_{jt} by observables z_{jt} is successful.

In Sections 4 and 5, we further develop the consequences of unobserved demand heterogeneity and changes in firm conduct for the DLW method. We first characterize the bias in the estimated markup resulting from $\delta_{jt} \neq 0$. Then we provide an empirical application to test for the effects of $\delta_{jt} \neq 0$.

4 Bias in estimated markup

If there are differences in demand or conduct across firms or time that cannot be fully controlled for by z_{jt}, then $\delta_{jt} \neq 0$ and equation (4) becomes

$$q_{jt} = q^*_{jt} + \varepsilon_{jt} = \phi(z_{jt}, \delta_{jt}) + \varepsilon_{jt}. \tag{8}$$

Regressing q_{jt} on observables z_{jt} in the first step of ACF, however, yields an estimate of the conditional expectation $E(q_{jt}|z_{jt}) = E(q^*_{jt}|z_{jt}) = E(\phi(z_{jt}, \delta_{jt})|z_{jt}) = \widetilde{\phi}(z_{jt})$, where the first equality uses that ε_{jt} is mean independent of z_{jt}. We thus write the first-stage regression as

$$q_{jt} = \widetilde{\phi}(z_{jt}) + \phi(z_{jt}, \delta_{jt}) - \widetilde{\phi}(z_{jt}) + \varepsilon_{jt} = \widetilde{\phi}(z_{jt}) + \zeta_{jt} + \varepsilon_{jt} = \widetilde{\phi}(z_{jt}) + \widetilde{\varepsilon}_{jt},$$

where the prediction error $\zeta_{jt} = q^*_{jt} - E(q^*_{jt}|z_{jt}) = \phi(z_{jt}, \delta_{jt}) - \widetilde{\phi}(z_{jt})$ is mean independent of z_{jt} by construction and has mean zero.

[13] While the firm's cost-minimizing decisions depend indirectly on δ_{jt} through Q^*_{jt}, its profit-maximizing decisions depend directly on δ_{jt} and thus cannot be used either to replace Hicks-neutral productivity ω_{jt} by observables z_{jt}. Moreover, under imperfect competition the profit-maximizing demand for a variable input may not be invertible even if $\delta_{jt} = 0$ (Biondi 2022). DLW do not state if the inverse in their equation (9) is derived from the profit-maximizing or cost-minimizing demand for materials.

We develop four alternative characterizations of the prediction error ζ_{jt} that are helpful in assessing the bias in the estimated disturbance in the first step of ACF and the estimated output elasticity in the second step. From the production function in equation (1), we have $q^*_{jt} = \phi(z_{jt}, \delta_{jt}) = \ln F(K_{jt}, L_{jt}, M_{jt}) + \omega_{jt}$. Assuming for simplicity that the demand the firm faces takes the form $Q^*_{jt} = D(z^D_{jt}) \exp(\delta_{jt})$, we also have $q^*_{jt} = \phi(z_{jt}, \delta_{jt}) = \ln D(z^D_{jt}) + \delta_{jt}$. Recalling that z_{jt} includes input quantities and prices and that $z^D_{jt} \subseteq z_{jt}$ (see Section 3) it follows that

$$\zeta_{jt} = \omega_{jt} - E(\omega_{jt}|z_{jt}) = \delta_{jt} - E(\delta_{jt}|z_{jt}). \tag{9}$$

Our first two characterizations in equation (9) show that ζ_{jt} covaries with any part of Hicks-neutral productivity ω_{jt} and any part of the demand shock δ_{jt} that is not captured by observables z_{jt}.

To develop our next two characterizations of ζ_{jt}, we use the FOC for variable input X_{jt} in equation (2) to write marginal cost MC_{jt} as

$$MC_{jt} = \frac{1}{P_{Xjt} \frac{\partial F(K_{jt}, L_{jt}, M_{jt})}{\partial X_{jt}}} \exp(-\omega_{jt}). \tag{10}$$

Using equation (9) it follows that

$$\ln MC_{jt} - E\left(\ln MC_{jt}|z_{jt}\right) = -\omega_{jt} + E\left(\omega_{jt}|z_{jt}\right) = -\zeta_{jt}. \tag{11}$$

Hence, $\zeta_{jt} = -\ln MC_{jt} + E(\ln MC_{jt}|z_{jt})$ is inversely related with any true determinant of marginal cost that has not been controlled for by observables z_{jt}. Finally, using equation (11) and assuming that the price of output P_{jt} is included in z_{jt} (as in De Loecker et al. 2016) it follows that

$$\ln \mu_{jt} - E(\ln \mu_{jt}|z_{jt}) = \zeta_{jt}. \tag{12}$$

Hence, $\zeta_{jt} = \ln \mu_{jt} - E(\ln \mu_{jt}|z_{jt})$ covaries with any true determinant of the markup that has not been controlled for by observables z_{jt}.

Bias in estimated disturbance. With $\delta_{jt} \neq 0$, DLW obtain an estimate of $\widetilde{\varepsilon}_{jt} = \zeta_{jt} + \varepsilon_{jt}$ in the first step of ACF and substitute this biased estimate $\widehat{\widetilde{\varepsilon}}_{jt}$ into equation (3) in lieu of ε_{jt}. Assuming for now that the output elasticity $\beta_X(\cdot)$ is known, DLW therefore obtain $\mu^{DLW}_{jt} = \mu_{jt} \exp(-\zeta_{jt})$ or, equivalently, $\ln \mu^{DLW}_{jt} = \ln \mu_{jt} - \zeta_{jt}$. Equation (12) implies that $\widetilde{\varepsilon}_{jt}$ covaries with any true determinant of the markup that has not been controlled for by observables z_{jt}. The markup μ_{jt}^{DLW} estimated by the DLW method is therefore inversely related with any true determinant of the markup that has not been controlled for by observables z_{jt}.

The economist in the opening paragraph in Section 1 who uses the DLW method to estimate the markup and then regresses $\ln \mu_{jt}^{DLW}$ on a step dummy that is one for the acquiring firms in the "after" period and zero otherwise is a case in point: the bias in $\widehat{\widehat{\varepsilon}}_{jt}$ predisposes her to finding that the markup of the acquiring firms has not increased or even decreased following the wave of acquisitions.

We provide an analytically tractable illustration of how severely biased the regression of the markup $\ln \mu_{jt}^{DLW}$ obtained by the DLW method on the step dummy may be. In Appendix A, we model a merger between two symmetric Bertrand competitors. The production function of firm j is $Q_{jt}^* = V_{jt}^{\beta_V} \exp(\omega_{jt})$, where $\beta_V \in (0,1)$ is the output elasticity of the variable input V_{jt}. We think of V_{jt} as a composite of labor and materials akin to cost of goods sold and, to simplify the exposition, we abstract from capital and specify $P_{Vjt} = 1$. The demand firm j faces is $Q_{jt}^* = \exp(\delta_{jt}) P_{jt}^{\eta} P_{-jt}^{\gamma}$, where $\eta < -1$ and $\gamma > \eta$ are the own- and cross-price elasticities of demand and δ_{jt} is a demand shock. As we show in Appendix A, the markup of firm j is $\mu_{jt} = \frac{\eta + \iota_{jt}\gamma}{\eta + \iota_{jt}\gamma + 1}$, where pre-merger $\iota_{jt} = 0$ and post-merger $\iota_{jt} = 1$. The wave of acquisitions may therefore enable a firm to increase its markup from $\underline{\mu} = \frac{\eta}{\eta+1}$ to $\overline{\mu} = \frac{\eta+\gamma}{\eta+\gamma+1}$.

In Appendix B, we use this model as the basis for the data generating process. For simplicity, we assume that mergers occur exogenously and use $\lambda = \Pr(\iota_{jt} = 1) \in (0,1)$ to denote the probability that firm j in period t charges the higher markup $\overline{\mu}$. We further assume that $\omega_{jt} \sim N(0, \sigma_\omega^2)$ and $\delta_{jt} \sim N(0, \sigma_\delta^2)$.[14]

Regressing the markup $\ln \mu_{jt}$ on the step dummy ι_{jt} yields the coefficient of interest $\ln \overline{\mu} - \ln \underline{\mu}$ that measures the increase in market power following the wave of acquisitions. In contrast, the regression of the markup $\ln \mu_{jt}^{DLW} = \ln \mu_{jt} - \zeta_{jt}$ obtained by the DLW method on the step dummy may be severely biased to the extent that the prediction error ζ_{jt} covaries with any part of the step dummy ι_{jt} that is not fully controlled for by observable z_{jt}. Because the

[14]If ω_{jt} follows an $AR(1)$ process with parameter ρ and innovation $\xi_{jt} \sim N\left(0, (1-\rho^2)\sigma_\omega^2\right)$, then $\omega_{jt} \sim N\left(0, \sigma_\omega^2\right)$ at the stationary distribution. Persistence may similarly be introduced into δ_{jt}.

price of output P_{jt} is informative about the degree of market power that a firm enjoys, we include it in the first step of ACF and specify $z_{jt} = (v_{jt}, p_{jt})$.

As we show in Appendix B, regressing $\ln \mu_{jt}^{DLW}$ on the step dummy yields

$$\left(\ln \overline{\mu} - \ln \underline{\mu}\right) \left(E(\tau(z_{jt})|\iota_{jt} = 1) - E(\tau(z_{jt})|\iota_{jt} = 0)\right) < \ln \overline{\mu} - \ln \underline{\mu}, \qquad (13)$$

where $\tau(z_{jt}) \in (0, 1)$ is a weight that depends on the probability $\lambda = \Pr(\iota_{jt} = 1)$ and the densities of z_{jt} conditional on $\iota_{jt} = 1$ and $\iota_{jt} = 0$, respectively (see equation (35) in Appendix B for $\tau(z_{jt})$ and equation (38) for $E(\tau(z_{jt})|\iota_{jt})$). It follows immediately from $\tau(z_{jt}) \in (0, 1)$ that the coefficient of interest in equation (13) is biased down. Moreover, $\tau(z_{jt}) = 0$ if $\lambda = 0$ and $\tau(z_{jt}) = 1$ if $\lambda = 1$ and the regression estimates the coefficient of interest to be zero in these extreme cases. By continuity, the coefficient of interest is therefore severely biased towards zero if either very few or almost all firms are able to increase their markup following the wave of acquisitions. In Appendix B, we evaluate the bias for intermediate cases. We also show that the coefficient of interest may be estimated to be negative instead of positive if the price of output P_{jt} is excluded from the first step of ACF.

Bias in estimated output elasticity. As the output elasticity $\beta_X(\cdot)$ is not known, it must be estimated in the second step of ACF. With $\delta_{jt} \neq 0$, $\phi(z_{jt})$ in equation (5) becomes $\phi(z_{jt}, \delta_{jt})$. DLW obtain an estimate of $\widetilde{\phi}(z_{jt})$ in the first step of ACF and substitute it into equation (5) in lieu of $\phi(z_{jt}, \delta_{jt})$. Rewriting equation (5) accordingly yields

$$\begin{aligned} q_{jt} &= \ln F(K_{jt}, L_{jt}, M_{jt}) \\ &\quad + g\left(\widetilde{\phi}(z_{jt-1}) + \zeta_{jt-1} - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})\right) + \xi_{jt} + \varepsilon_{jt}. \end{aligned} \qquad (14)$$

To take the most favorable case, let ω_{jt} follow an $AR(1)$ process with parameter ρ so that $g(\omega_{jt-1}) = \rho\omega_{jt-1}$ and $\rho\zeta_{jt-1}$ becomes part of the composite error term. Even in this case, however, $\delta_{jt} \neq 0$ invalidates commonly used instruments in the second step of ACF.

Current capital K_{jt} is commonly used as an instrument in the second step of ACF. The underlying assumption is that the firm decides on investment, and thus capital, in period $t-1$ before it observes Hicks-neutral productivity ω_{jt} ("time to build"); hence, K_{jt} is uncorrelated with the productivity innovation ξ_{jt}. Under an analogous assumption, current labor L_{jt}

is also often used as an instrument (De Loecker et al. 2016, p. 471). However, none of these instruments is valid with $\delta_{jt} \neq 0$: any variable that is chosen in period $t-1$ is chosen with knowledge of Hicks-neutral productivity ω_{jt-1} and the demand shock δ_{jt-1} and, from equation (9), is therefore correlated with what remains of ω_{jt-1} and δ_{jt-1} in the prediction error $\zeta_{jt-1} = \omega_{jt-1} - E(\omega_{jt-1}|z_{jt-1}) = \delta_{jt-1} - E(\delta_{jt-1}|z_{jt-1})$ after controlling for z_{jt-1}.[15] Using invalid instruments biases the estimated output elasticity.

To formalize this bias, we extend the production function of firm j to be $Q^*_{jt} = K_{jt}^{\beta_K} V_{jt}^{\beta_V} \exp(\omega_{jt})$. Capital K_{jt} is chosen in period $t-1$ and the variable input V_{jt} is chosen in period t. We specify $z_{jt} = (k_{jt}, v_{jt}, p_{Vjt})$ in the first step of ACF and estimate $E(q^*_{jt}|z_{jt}) = \widetilde{\phi}(z_{jt})$. We simplify the second step of ACF by assuming that ρ is known. We accordingly rewrite equation (14) as

$$q_{jt} - \rho\widetilde{\phi}(z_{jt-1}) = \beta_K (k_{jt} - \rho k_{jt-1}) + \beta_V (v_{jt} - \rho v_{jt-1}) + \rho\zeta_{jt-1} + \xi_{jt} + \varepsilon_{jt} \tag{15}$$

and use 2SLS with instruments k_{jt}, p_{Vjt}, and z_{jt-1} to estimate β_K and β_V. Using $k_{jt} - \rho k_{jt-1}$ and $\widehat{v}_{jt} - \rho v_{jt-1}$, where $\widehat{v}_{jt} = E(v_{jt}|k_{jt}, p_{Vjt}, z_{jt-1})$, to denote the projections of the regressors $k_{jt} - \rho k_{jt-1}$ and $v_{jt} - \rho v_{jt-1}$ on the instruments, the 2SLS estimator converges to

$$\begin{aligned}
&\begin{pmatrix} \beta_K \\ \beta_V \end{pmatrix} + \begin{pmatrix} E((k_{jt} - \rho k_{jt-1})^2) & E((k_{jt} - \rho k_{jt-1})(v_{jt} - \rho v_{jt-1})) \\ E((k_{jt} - \rho k_{jt-1})(\widehat{v}_{jt} - \rho v_{jt-1})) & E((\widehat{v}_{jt} - \rho v_{jt-1})(v_{jt} - \rho v_{jt-1})) \end{pmatrix}^{-1} \\
&\qquad \cdot \begin{pmatrix} E((k_{jt} - \rho k_{jt-1})(\rho\zeta_{jt-1} + \xi_{jt} + \varepsilon_{jt})) \\ E((\widehat{v}_{jt} - \rho v_{jt-1})(\rho\zeta_{jt-1} + \xi_{jt} + \varepsilon_{jt})) \end{pmatrix} \\
&= \begin{pmatrix} \beta_K \\ \beta_V \end{pmatrix} + \rho \begin{pmatrix} E((k_{jt} - \rho k_{jt-1})^2) & E((k - \rho k_{jt-1})(v_{jt} - \rho v_{jt-1})) \\ E((k_{jt} - \rho k_{jt-1})(v_{jt} - \rho v_{jt-1})) & E((\widehat{v}_{jt} - \rho v_{jt-1})(v_{jt} - \rho v_{jt-1})) \end{pmatrix}^{-1} \\
&\qquad \cdot \begin{pmatrix} E(k_{jt}\zeta_{jt-1}) \\ E(\widehat{v}_{jt}\zeta_{jt-1}) \end{pmatrix},
\end{aligned} \tag{16}$$

where the equality uses that (i) the instruments are uncorrelated with ξ_{jt} and ε_{jt} by assumption, (ii) $E((k_{jt} - \rho k_{jt-1})(\widehat{v}_{jt} - \rho v_{jt-1})) = E((k_{jt} - \rho k_{jt-1})(\widehat{v}_{jt} - v_{jt} + v_{jt} - \rho v_{jt-1})) = E((k_{jt} - \rho k_{jt-1})(v_{jt} - \rho v_{jt-1}))$ because $\widehat{v}_{jt} - v_{jt}$ is mean independent of k_{jt} and k_{jt-1} by construction of $\widehat{v}_{jt}$, and (iii) the prediction error ζ_{jt-1} is mean independent of k_{jt-1} and v_{jt-1} by virtue of the first step of ACF.

[15] Recall that z_{jt} is specified from the arguments of the demand that is inverted for ω_{jt} (see Section 3).

Equation (4) makes clear that $E(k_{jt}\zeta_{jt-1}) \neq 0$ generally implies the existence of a bias. In what follows, we let $bias = \frac{\widehat{\beta}_V - \beta_V}{\beta_V}$ denote the relative bias in the estimate of the output elasticity of the variable input V_{jt}. Assessing this bias requires assessing the various terms in equation (16). This, in turn, requires specifying the data generating process. In Appendix C, we do so in a way that allows us to analytically characterize $bias = \frac{\widehat{\beta}_V - \beta_V}{\beta_V}$.

Specifically, as we detail in Appendix C, marginal cost $MC(K_{jt}, P_{Vjt}, Q^*_{jt}, \omega_{jt})$ depends on the index $\frac{1}{\beta_V}\left(\beta_K k_{jt} - \beta_V p_{Vjt} + \omega_{jt}\right)$. Rather than specifying demand and assuming firm conduct, we start from marginal revenue $MR(Q^*_{jt}, \delta_{jt})$, where δ_{jt} is a demand shock, and model planned output as

$$q^*_{jt} = \frac{1}{\beta_V}\left(\beta_K k_{jt} - \beta_V p_{Vjt} + \omega_{jt}\right) + \delta_{jt}. \tag{17}$$

Equation (17) may be viewed as a log-linearization of the solution to $MR(Q^*_{jt}, \delta_{jt}) = MC(K_{jt}, P_{Vjt}, Q^*_{jt}, \omega_{jt})$.[16]

We further model the law of motion for capital as

$$k_{jt+1} = \tau_0 + \tau_k k_{jt} + \tau_{p_K} p_{Kjt} + \tau_{p_V} p_{Vjt} + \tau_\omega \omega_{jt} + \tau_\delta \delta_{jt}, \tag{18}$$

where the price of capital P_{Kjt} may include, besides the user cost of capital, any other influences on capital such as investment opportunities and financial constraints. Equation (18) may be viewed as a log-linearization of the demand for investment that arises from a dynamic profit maximization problem (as in OP, see Section 3).[17] Alternatively, equation (18) encompasses as a special case the solution to a dynamic cost minimization problem in which capital is subject to time to build but not to adjustment costs, and we provide the resulting expressions for τ_0, τ_k, τ_{p_K}, τ_{p_V}, τ_ω, and τ_δ in equation (42) in Appendix C.

---

[16] If the demand firm j faces is $Q^*_{jt} = \exp(\delta_{jt}) P^\eta_{jt}$, where $\eta < -1$, then profit maximization implies

$$q^*_{jt} = \frac{\beta_V \eta}{\eta(1-\beta_V) - \beta_V}\left(\ln\frac{\beta_V(1+\eta)}{\eta} - \frac{\delta_{jt}}{\eta} + \frac{1}{\beta_V}\left(\beta_K k_{jt} - \beta_V p_{Vjt} + \omega_{jt}\right)\right).$$

As $\frac{\eta\beta_V}{\eta(1-\beta_V)-\beta_V} > 0$, q^*_{jt} increases in the index $\frac{1}{\beta_V}\left(\beta_K k_{jt} - \beta_V p_{Vjt} + \omega_{jt}\right)$ and in δ_{jt} in accordance with equation (17).

[17] Assuming constant returns to scale in production and quadratic adjustment costs to capital, the firm's dynamic programming problem can be solved in closed form if the firm is a price-taker in the output market (Syverson 2001, Van Biesebroeck 2007, Ackerberg et al. 2015). Unfortunately, this is no longer possible under imperfect competition.

Finally, we assume that $x_{jt} \in \{p_{Kjt}, p_{Vjt}, \omega_{jt}, \delta_{jt}\}$ follows the $AR(1)$ process $x_{jt} = \rho_x x_{jt-1} + \xi_{xjt}$ with parameter ρ_x and innovation $\xi_{xjt} \sim N(0, \sigma_x^2)$. To simplify the exposition, we omit the subscript ω from ρ_ω and $\xi_{\omega jt}$ in what follows.

In Appendix C, we show that there is no bias in three special cases:

Proposition 1 *If $\sigma_\delta^2 = 0$, then $\zeta_{jt} = 0$ and hence bias $= 0$.*

Proposition 1 rules out any differences in demand and conduct across firms and time as required by DLW.

Proposition 2 *If $\sigma_\omega^2 = 0$ or $\rho = 0$, then $E(k_{jt}\zeta_{jt-1}) = E(\widehat{v}_{jt}\zeta_{jt-1}) = 0$ and hence bias $= 0$.*

In stark contrast to the literature on productivity analysis, Proposition 2 rules out dispersion or persistence in productivity.

Proposition 3 *If $\tau_\omega = \tau_\delta = 0$, then $E(k_{jt}\zeta_{jt-1}) = E(\widehat{v}_{jt}\zeta_{jt-1}) = 0$ and hence bias $= 0$.*

Note that neither $\tau_\omega = 0$ nor $\tau_\delta = 0$ by itself is sufficient for $E(k_{jt}\zeta_{jt-1}) = E(\widehat{v}_{jt}\zeta_{jt-1}) = 0$. A bias may therefore arise even if the evolution of capital depends solely on either productivity or the demand shock. Of course, ω_{jt} is a state variable in the firm's dynamic programming problem whenever $\rho > 0$, as is δ_{jt} whenever $\rho_\delta > 0$, so that typically $\tau_\omega \neq 0$ and $\tau_\delta \neq 0$ (as in the above-mentioned dynamic cost minimization problem, see equation (42) in Appendix C).

Proposition 1 differs from Propositions 2 and 3 in that there is no prediction error in case of Proposition 1, whereas there is a prediction error in case of Propositions 2 and 3 but it is not transmitted to the estimated output elasticity.

Taken together, Propositions 1 and 3 show that current capital K_{jt} is no longer a valid instrument with $\delta_{jt} \neq 0$ to the extent that it is chosen in period $t-1$ with knowledge of Hicks-neutral productivity ω_{jt-1} and the demand shock δ_{jt-1} and is therefore correlated with what remains of ω_{jt-1} and δ_{jt-1} in the prediction error $\zeta_{jt-1} = \omega_{jt-1} - E(\omega_{jt-1}|z_{jt-1}) = \delta_{jt-1} - E(\delta_{jt-1}|z_{jt-1})$ after controlling for z_{jt-1}. The evolution of capital in response to Hicks-neutral productivity and the demand shock ensures that the prediction error in the first step of ACF gives rise to a bias in the estimated output elasticity in the second step of ACF.

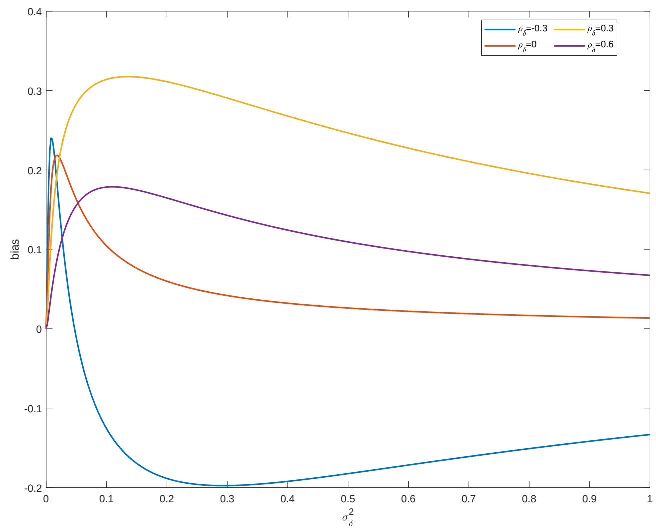

Figure 1: $bias = \frac{\widehat{\beta}_V - \beta_V}{\beta_V}$ over range of σ_δ^2 for $\rho_\delta \in \{-0.3, 0, 0.3, 0.6\}$. τ_0, τ_k, τ_{p_K}, τ_{p_V}, τ_ω, and τ_δ from dynamic cost minimization problem.

In evaluating the bias outside the special cases in Propositions 1–3, we focus on the role of the dispersion and persistence in the demand shock. We parameterize

$$\beta_K = 0.2, \quad \beta_V = 0.8, \quad \rho = 0.9, \quad \sigma_\omega^2 = 0.25\left(1 - 0.9^2\right),$$
$$\rho_\delta \in \{-0.3, 0, 0.3, 0.6\}, \quad \sigma_\delta^2 \in [0, 1],$$
$$\rho_{p_K} = \rho_{p_V} = 0, \quad \sigma_{p_K}^2 = \sigma_{p_V}^2 = 0.01$$

and determine τ_0, τ_k, τ_{p_K}, τ_{p_V}, τ_ω, and τ_δ in the law of motion for capital in equation (18) from the above-mentioned dynamic cost minimization problem (see equation (42) in Appendix C). Our parameterization implies $Var(\omega) = 0.25$ and is therefore broadly consistent with the dispersion and persistence in total factor productivity reported in the literature on productivity analysis. Our parameterization further approximates the widely used assumption in the literature following OP that input prices are constant. Figure 1 shows $bias = \frac{\widehat{\beta}_V - \beta_V}{\beta_V}$ over the specified range of σ_δ^2 for the various values of ρ_δ. As can be seen, the bias is small near $\sigma_\delta^2 = 0$ in line with Proposition 1 but rapidly becomes large. The bias may be positive or negative. In Appendix C, we show that the bias may be enormous if we change the parameterization of the law of motion for capital in equation (18).

Bias in estimated markup. Substituting biased estimates of the disturbance ε_{jt} and the output elasticity $\beta_X(\cdot)$ into equation (3), DLW obtain

$$\begin{aligned}\mu_{jt}^{DLW} &= \frac{\beta_X(K_{jt}, L_{jt}, M_{jt})(1 + bias_{jt})}{S_{Xjt}^R} \exp(-\varepsilon_{jt} - \widetilde{\varepsilon}_{jt} + \varepsilon_{jt}) \\ &= \mu_{jt}(1 + bias_{jt}) \exp(-\zeta_{jt}),\end{aligned}$$

where we index the bias by j and t to accommodate production functions other than the Cobb-Douglas from our example. It follows that

$$\ln \mu_{jt}^{DLW} \approx \ln \mu_{jt} + bias_{jt} - \zeta_{jt}.$$

The bias in the markup μ_{jt}^{DLW} obtained by the DLW method has two components. The first component affects the unconditional expectation of μ_{jt}^{DLW} and hence its level since

$$E(\ln \mu_{jt}^{DLW}) = E(\ln \mu_{jt}) + E(bias_{jt}). \tag{19}$$

The second component of the bias affects the conditional expectation of μ_{jt}^{DLW} and hence how it correlates with variables that the economist may be interested in such as a firm's export status or measures of trade liberalization. To see this, note that for any such variable w_{jt} we have

$$E(\ln \mu_{jt}^{DLW} | w_{jt}) = E(\ln \mu_{jt} | w_{jt}) + E(bias_{jt} | w_{jt}) - E(\zeta_{jt} | w_{jt}). \tag{20}$$

5 Testing for the effects of unobserved demand heteroegeneity

In this section, we test whether the first step of ACF is correctly specified as $q_{jt} = \phi(z_{jt}) + \varepsilon_{jt}$ (see again equation (4)) or becomes $q_{jt} = \phi(z_{jt}, \delta_{jt}) + \varepsilon_{jt}$ (equation (8)). The difficulty is that δ_{jt} is inherently unobservable, as is its correlation with the observables z_{jt}. We overcome this difficulty by exploiting that our data contains a firm- and year-specific assessment of the evolution of a firm's main market (slump, stability, or expansion). As the underlying survey question intends to measure changes in market size, this market dynamism variable mdy_{jt} is as good a proxy for shifts in the demand a firm faces as one can hope for in production data and therefore an important component of δ_{jt}. At the same time, there is no reason to believe that it captures differences in demand across firms or time in their entirety. Hence, while our market dynamism variable mdy_{jt} is useful for testing purposes, adding it to the observables z_{jt} in the first step of ACF is not a solution to the problem of unobserved demand heterogeneity.

Data. Our data come from the Encuesta Sobre Estrategias Empresariales (ESEE) survey, a firm-level survey of the Spanish manufacturing sector, and spans 1990-2012. An attractive feature of our data is that it contains firm- and year-specific Paasche-type price indices for output and materials that we use to deflate revenue and the value of materials.[18] Appendix D provides details on the sample and variables. We estimate the production function separately for 10 industries.

Specification and estimation. We specify a Cobb-Douglas production function $\ln F(K_{jt}, L_{jt}, M_{jt}) = \beta_0 + \beta_t + \beta_K k_{jt} + \beta_L l_{jt} + \beta_M m_{jt}$, where β_0 is a constant and β_t is a set of 21 year dummies. As in ACF, we specify an $AR(1)$ process with parameter ρ for Hicks-neutral productivity ω_{jt}.

We invert the demand for a variable input and write $\omega_{jt} = h(z_{jt})$. We include input quantities k_{jt}, l_{jt}, and m_{jt}, the real price of labor $p_{Ljt} - p_{jt}$, and the real price of materials $p_{Mjt} - p_{jt}$ in the observables z_{jt} in addition to the constant and the year dummies.

In the first step of ACF, we flexibly approximate $\phi(z_{jt})$ in equation (4) by a complete polynomial of order 3 in the continuous variables included in z_{jt}, the constant, and the year dummies and estimate by OLS. In the second step of ACF, we estimate equation (5) by GMM. The instruments are k_{jt}, k_{jt-1}, l_{jt-1}, m_{jt-1}, and $\widehat{\phi}(z_{jt-1})$ in addition to the constant and the year dummies. We correct the standard errors for the two-step nature of the estimation (see Appendix E).

[18] While using firm-specific price indexes is widely considered preferable to using industry-wide price indexes or estimating a revenue production function, well-known issues remain regarding comparing units of differentiated products across firms and multi-product production within firms (De Loecker & Syverson 2021, pp. 175–182).

Markup. While our approach extends directly to alternative assumptions, we assume that both labor L_{jt} and materials M_{jt} are variable inputs.[19] Combining equation (3) for labor and materials yields

$$\mu_{jt} = \frac{\nu(K_{jt}, L_{jt}, M_{jt})}{S^R_{Ljt} + S^R_{Mjt}} \exp(-\varepsilon_{jt}), \tag{21}$$

where $\nu(\cdot) = \beta_L(\cdot) + \beta_M(\cdot) = \frac{\partial \ln F(\cdot)}{\partial l_{jt}} + \frac{\partial \ln F(\cdot)}{\partial m_{jt}}$ is the short-run elasticity of scale.[20] As in DLW, we obtain the markup μ^{DLW}_{jt} of firm j in period t by substituting estimates of the parameters $\nu = \beta_L + \beta_M$ of the Cobb-Douglas production function and of the disturbance ε_{jt} into equation (21).

Results. Table 1 reports the results from the DLW method. Column (1) shows the average (log) markup by industry, along with the sample standard deviation. The average markup ranges from 0.090 in industry 1 to 0.445 in industry 3.

Columns (2)–(4) show the underlying production function estimates. The estimate of the output elasticity of capital β_K is plausible although not significant at the 5% level in industries 5, 6, and 8. The estimate of the short-run elasticity of scale ν is on the high side and in 7 industries ranges from 0.956 to 1.173.

While the extant literature does not routinely conduct formal specification tests, the Sargan test in column (5) rejects the specification in 3 industries at the 5% significance level. This is not surprising: as shown in Section 4, if there are differences in demand across firms or time that cannot be fully controlled for by z_{jt}, then $\delta_{jt} \neq 0$ and k_{jt} is no longer a valid instrument in the second step of ACF.

Test. To more specifically test for the effects of unobserved demand heterogeneity, recall from equation (14) that if $\delta_{jt} \neq 0$, then equation (5) in the second step of ACF becomes

$$q_{jt} = \ln F(K_{jt}, L_{jt}, M_{jt})$$

[19] See again footnote 4.

[20] Doraszelski & Jaumandreu (2019) and Raval (2022) show that, in practice, the level of the estimated markup and its correlation with variables of interest can be different depending on whether labor or materials is used in equation (3). Combining them in equation (21) ameliorates this problem.

Table 1: DLW method

	Baseline specification				Sargan test		Incl. mdy	Regression on mdy[a]	
	Markup A (s. dev.)	β_K (s. e.)	ν (s. e.)	ρ (s. e.)	p-val. (1 d. f.)	p-val. (2 d. f.)	Markup B (s. dev.)	Markup A (s. e.)	Markup B (s. e.)
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1. Metals and metal products	0.090	0.087	0.886	0.792	0.214	0.012	0.104	-0.017*	0.036*
	(0.091)	(0.019)	(0.025)	(0.033)			(0.095)	(0.006)	(0.007)
2. Non-metallic minerals	0.195	0.066	0.964	0.505	0.030	0.013	0.202	-0.003	0.067*
	(0.119)	(0.023)	(0.030)	(0.023)			(0.125)	(0.012)	(0.015)
3. Chemical products	0.445	0.061	1.173	0.953	0.225	0.044	0.246	0.015	0.060*
	(0.166)	(0.029)	(0.117)	(0.102)			(0.169)	(0.017)	(0.018)
4. Agric. and ind. machinery	0.347	0.060	1.102	0.928	0.226	0.005	0.198	-0.008	0.024*
	(0.108)	(0.028)	(0.073)	(0.106)			(0.109)	(0.011)	(0.011)
5. Electrical goods	0.364	0.021	1.089	0.918	0.655	0.023	0.251	-0.003	0.035*
	(0.156)	(0.025)	(0.109)	(0.165)			(0.157)	(0.017)	(0.018)
6. Transport equipment	0.118	0.041	0.917	0.824	0.815	0.248	0.145	0.014	0.061*
	(0.130)	(0.031)	(0.051)	(0.060)			(0.133)	(0.014)	(0.015)
7. Food, drink and tobacco	0.158	0.080	0.905	0.829	0.001	0.000	0.157	-0.067*	-0.040*
	(0.171)	(0.021)	(0.027)	(0.028)			(0.172)	(0.014)	(0.014)
8. Textile, leather and shoes	0.142	0.035	0.965	0.840	0.548	0.000	0.185	0.001	0.049*
	(0.092)	(0.021)	(0.027)	(0.027)			(0.096)	(0.007)	(0.008)
9. Timber and furniture	0.169	0.120	0.988	0.922	0.002	0.001	0.213	-0.012	0.045*
	(0.102)	(0.031)	(0.076)	(0.074)			(0.107)	(0.010)	(0.011)
10. Paper and printing products	0.239	0.072	0.956	0.820	0.131	0.003	0.244	-0.032	0.024
	(0.150)	(0.019)	(0.029)	(0.039)			(0.152)	(0.017)	(0.017)

[a] An asterisk indicates that the coefficient is significant at the 5% level.

$$+\rho\left(\widetilde{\phi}(z_{jt-1}) - \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1})\right) + \rho\zeta_{jt-1} + \xi_{jt} + \varepsilon_{jt} \quad (22)$$

where $\widetilde{\phi}(z_{jt}) = E\left(\phi(z_{jt}, \delta_{jt})|z_{jt}\right)$ and $\zeta_{jt} = \phi(z_{jt}, \delta_{jt}) - \widetilde{\phi}(z_{jt})$ is the prediction error. Treating our market dynamism variable mdy_{jt} as a component of δ_{jt} and noting that $\delta_{jt} \neq 0$ generally implies $\zeta_{jt} \neq 0$, we test for $\delta_{jt} \neq 0$ by examining the correlation of mdy_{jt} with the composite error $\rho\zeta_{jt-1} + \xi_{jt} + \varepsilon_{jt}$. Adding mdy_{jt} to the instruments used in the second step of ACF, the Sargan test in column (6) detects a significant correlation and rejects the specification in 9 industries at the 5% significance level.

While the Sargan test points to unobserved demand heterogeneity, it may be difficult to detect this problem from a routine examination of the average markup or the coefficient of determination in the first step of ACF. Indeed, the R^2 exceeds 0.99 in all industries.

Bias in estimated markup: level component. As shown in Section 4, unobserved demand heterogeneity causes a bias in the estimate of the disturbance ε_{jt} and a bias in the estimate of the output elasticity $\beta_X(\cdot)$. Plugging biased estimates into equation (21), in turn, causes a bias in the estimated markup μ_{jt}^{DLW} that has two components. The first component affects the level of the estimated markup and the second component how it correlates with variables of interest (see again equations (19) and (20)). We illustrate both components in turn.

Starting with the level component, we include mdy_{jt} in z_{jt}, re-estimate equations (4) and (5), and re-compute the markup μ_{jt}^{DLW}.[21] Column (7) of Table 1 shows the average (log) markup by industry. Because mdy_{jt} is only a proxy for δ_{jt}, there is no reason to believe that the estimates are entirely free of bias. Nevertheless, including mdy_{jt} in z_{jt} decreases the markup noticeably in industries 3, 4, and 5 compared to the baseline in column (1).

As shown in Section 4, if $\delta_{jt} \neq 0$, then k_{jt} is no longer a valid instrument in the second step of ACF. A natural response may be for the researcher to drop this instrument. If the data has been generated by an $AR(1)$ process for Hicks-neutral productivity ω_{jt}, then this way of proceeding, in theory, avoids the level component of the bias, although it does not avoid the correlation

[21] We maintain the assumption that ε_{jt} is mean independent of z_{jt}. We also include mdy_{jt} as an instrument in the second step of ACF. We proceed similarly below with the indicator of the firm's export status xst_{jt}.

component. How well it works in practice no doubt depends on the data at hand.

Bias in estimated markup: correlation component. Turning to the correlation component, we regress $\ln \mu_{jt}^{DLW}$ on our market dynamism variable mdy_{jt} and report the estimated coefficient in Table 1.[22] In the baseline with mdy_{jt} excluded from z_{jt}, the estimated markup is not significantly correlated with market dynamism in 8 industries and significantly negatively correlated with market dynamism in 2 industries (column (8)). In contrast, with mdy_{jt} included in z_{jt}, the estimated markup is significantly positively correlated in 8 industries (column (9)). The latter conveys, as expected, that firms enjoy a higher markup if their demands are expanding rather than contracting.

This reversal happens because, as shown in Section 4, with mdy_{jt} excluded from z_{jt}, a large part of demand heterogeneity is left in the estimated disturbance. This estimate of $\widetilde{\varepsilon}_{jt} = \zeta_{jt} + \varepsilon_{jt}$ is, in turn, substituted into equation (21) in lieu of ε_{jt} to obtain the markup μ_{jt}^{DLW}. Including mdy_{jt} in z_{jt} absorbs an additional part of demand heterogeneity. The resulting change in the estimated disturbance rectifies the correlation of the estimated markup with market dynamism.

It turns out that if the data is generated by a Cobb-Douglas production function and if the researcher is only interested in the correlation of the estimated markup with a variable of interest such as mdy_{jt}, then she may proceed by simply including this variable in the first step of ACF. This amounts to purpose-building the markup for the ex-post analysis and does not address the level component of the bias.

To see why including mdy_{jt} in z_{jt} suffices to consistently estimate the correlation of the markup with market dynamism, consider equation (20). Because the short-run elasticity of scale ν is a constant for a Cobb-Douglas production function, so is $bias_{jt}$, and thus $E(bias_{jt}|mdy_{jt})$ is a constant that is absorbed into the constant of the regression. Moreover, if mdy_{jt} is included in z_{jt}, then $E(\zeta_{jt}|mdy_{jt}) = 0$. Therefore, $E\left(\ln \mu_{jt}^{DLW}|mdy_{jt}\right) = E\left(\ln \mu_{jt}|mdy_{jt}\right) + const.$[23]

[22] We include a constant and a set of 21 year dummies in this and all subsequent regressions of this type.

[23] Instead of including mdy_{jt} in z_{jt}, one can rewrite equation (21) as $\ln \mu_{jt} + \varepsilon_{jt} = \ln \nu(K_{jt}, L_{jt}, M_{jt}) - \ln\left(S_{Ljt}^R + S_{Mjt}^R\right)$. If $E(bias_{jt}|mdy_{jt})$ is a constant, then assuming that ε_{jt} is mean independent of mdy_{jt} and regressing $\ln \mu_{jt} + \varepsilon_{jt}$ on mdy_{jt} suffices to

Of course, if a Cobb-Douglas production function is not appropriate for the data at hand, then $E(bias_{jt}|mdy_{jt})$ is generally not a constant and thus cannot be absorbed into the constant of the regression of $\ln \mu_{jt}^{DLW}$ on mdy_{jt}. It follows that the regression cannot consistently estimate the correlation of the markup with market dynamism even if mdy_{jt} is included in z_{jt} to ensure $E(\zeta_{jt}|mdy_{jt}) = 0$.

DLW focus on the correlation of the estimated markup with a firm's export status and accordingly include the firm's export status and "other demand conditions" in z_{jt} (p. 2446). They find "that exporters charge, on average, higher markups and that markups increase upon export entry" (p. 2437). This finding is at variance with a number of papers that have found similar or lower markups for the same products in the more competitive export markets (Bernstein & Mohnen 1991, Bughin 1996, Moreno & Rodriguez 2004, Das, Roberts & Tybout 2007, Jaumandreu & Yin 2018, Blum, Claro, Horstmann & Rivers 2023).

We only partially replicate the finding in DLW. In a first pass, we revert to the estimated markups in columns (1) and (7) of Table 1 and regress $\ln \mu_{jt}^{DLW}$ on an indicator of the firm's export status xst_{jt}. As can be seen in columns (1) and (2) of Table 2, the estimated coefficient on xst_{jt} is significantly positive in 6 industries and significantly negative in 1 industry. In a second pass, we re-estimate equations (4) and (5) and re-compute the markup μ_{jt}^{DLW} while including both our market dynamism variable mdy_{jt} and the indicator of the firm's export status xst_{jt} in z_{jt}. The estimated coefficient on xst_{jt} in the regression of $\ln \mu_{jt}^{DLW}$ on xst_{jt} is significantly positive in 5 industries and significantly negative in 2 industries, as can be seen in column (3) of Table 2. This illustrates that purpose-building the markup for the ex-post analysis may not be a straightforward undertaking.

6 Controlling for marginal cost

To consistently estimate markups, the DLW method has to either rule out any differences in demand and conduct across firms and time or assume that they can be fully controlled for by observables z_{jt}, as shown in Sections 3 and 4. As shown in Section 5, the latter may be difficult in practice even if proxies for shifts in demand such as our market dynamic variable mdy_{jt} and the indicator of the firm's export status xst_{jt} are available. De Loecker et al.

consistently estimate the correlation of the markup with market dynamism.

Table 2: DLW method (cont'd)

	Regression on xst^a		
	Markup A (s. e.)	Markup B (s. e.)	Markup C (s. e.)
	(1)	(2)	(3)
1. Metals and metal products	-0.007	-0.007	-0.002
	(0.008)	(0.008)	(0.008)
2. Non-metallic minerals	0.042*	0.045*	0.061*
	(0.012)	(0.013)	(0.012)
3. Chemical products	0.046*	0.045*	0.049*
	(0.015)	(0.015)	(0.016)
4. Agric. and ind. machinery	0.037*	0.038*	0.042*
	(0.013)	(0.013)	(0.013)
5. Electrical goods	0.033*	0.032*	0.028
	(0.015)	(0.016)	(0.016)
6. Transport equipment	0.001	0.006	0.013
	(0.022)	(0.022)	(0.022)
7. Food, drink and tobacco	-0.053*	-0.051*	-0.035*
	(0.015)	(0.015)	(0.016)
8. Textile, leather and shoes	0.027*	0.026*	0.036*
	(0.008)	(0.009)	(0.009)
9. Timber and furniture	0.033*	0.033*	0.049*
	(0.010)	(0.010)	(0.010)
10. Paper and printing products	-0.029	-0.029	-0.044*
	(0.016)	(0.016)	(0.017)

[a] An asterisk indicates that the coefficient is significant at the 5% level.

(2016) more broadly include variables such as location, product dummies, export status, input and output tariffs, market share, and the price of output. In this section, we provide a way to formally assess whether the endeavor of controlling for planned output Q^*_{jt} by observables z_{jt} is successful, so that the assumption $\delta_{jt} = 0$ required by DLW is justified.

A key insight from Sections 3 and 4 is that controlling for planned output Q^*_{jt} is equivalent to controlling for marginal cost $MC(K_{jt}, P_{Ljt}, P_{Mjt}, Q^*_{jt}, \omega_{jt}) = MC_{jt}$. Equation (11) formalizes that if marginal cost can be controlled for by observables z_{jt}, then $\zeta_{jt} = -\ln MC_{jt} + E(\ln MC_{jt}|z_{jt}) = 0$ and there is no prediction error in the first step of ACF. Equation (7), however, makes clear that marginal cost cannot be inferred from the assumption of cost minimization. Controlling for marginal cost therefore requires additional assumptions.

Given demand and cost, firm conduct is specified by a model of product market competition. To control for marginal cost we exploit that many such models entail the assumption of short-run profit maximization. Short-run profit maximization, in turn, implies that planned output Q^*_{jt} is determined by $MR(Q^*_{jt}, \delta_{jt}) = MC(K_{jt}, P_{Ljt}, P_{Mjt}, Q^*_{jt}, \omega_{jt})$, where $MR(\cdot)$ is marginal revenue and the demand shock δ_{jt} captures unobserved demand heterogeneity across firms and time and may include rivals' prices or planned quantities to the extent that these variables are unobserved in production data. Assuming short-run profit maximization therefore allows controlling for marginal cost by controlling for marginal revenue.[24] We illustrate the practical implication of this observation through a series of examples.

Example 1: Monopoly, Bertrand competition, or monopolistic competition with CES demand. The demand firm j faces is

$$Q^*_{jt} = \exp(\delta_{1jt}) P_{jt}^{\delta_{2jt}},$$

where $\delta_{2jt} < -1$ is the price elasticity. δ_{1jt} parameterizes market size in case of a monopoly. In case of Bertrand competition, δ_{1jt} is a function of rivals' prices. In case of monopolistic competition, δ_{1jt} is an aggregate statistic that characterizes the distribution of prices in the industry. As the impact of firm j on this statistic is assumed to be negligible in this case, firm j behaves as a monopoly given the demand it faces.

[24]Without the assumption of short-run profit maximization the state variables in the firm's dynamic programming problem must in general also be controlled for.

The marginal revenue of firm j is

$$MR(Q_{jt}^*, \delta_{jt}) = \frac{1+\delta_{2jt}}{\delta_{2jt}} \left(\exp(-\delta_{1jt}) Q_{jt}^*\right)^{\frac{1}{\delta_{2jt}}} = \frac{1+\delta_{2jt}}{\delta_{2jt}} P_{jt},$$

where $\delta_{jt} = (\delta_{1jt}, \delta_{2jt})$ is the (two-dimensional) demand shock. Consequently, if $\delta_{2jt} = const$, then the price of output P_{jt} controls for $MR(\cdot)$ and hence for $MC(\cdot)$. The simple expedient of including P_{jt} in z_{jt} in the first step of ACF ensures $\zeta_{jt} = 0$.

Note, however, that $\delta_{2jt} = const$ implies that the markup is $\mu_{jt} = \frac{\delta_{2jt}}{1+\delta_{2jt}} = const$. This is difficult to reconcile with the fact that the estimated markup from the DLW method varies across firms and time.

If $\delta_{2jt} \neq const$, then the price of output P_{jt} no longer suffices to control for $MR(\cdot)$ and including it in z_{jt} no longer suffices to ensure $\zeta_{jt} = 0$. Similarly, including P_{jt} in z_{jt} no longer suffices to ensure $\zeta_{jt} = 0$ if there are unobserved changes in firm conduct, e.g., following a wave of acquisitions.

Example 2: Bertrand competition with logit demand. The demand firm j faces is

$$Q_{jt}^* = \exp(\delta_{Mjt}) \frac{\exp(\delta_{Pjt} P_{jt})}{1 + \sum_{j=1}^{N_t} \exp(\delta_{Pjt} P_{jt})} = \frac{\exp(\delta_{Mjt})}{1 + \exp(\delta_{Ojt} - \delta_{Pjt} P_{jt})},$$

where $\delta_{Pjt} < 0$ parameterizes price sensitivity, δ_{Mjt} market size, and $\delta_{Ojt} = \ln\left(1 + \sum_{-j} \exp\left(\delta_{Pjt} P_{-jt}\right)\right)$ the combined impact of the outside good and the prices of the $N_t - 1$ rivals of firm j.

The marginal revenue of firm j is

$$\begin{aligned} MR(Q_{jt}^*, \delta_{jt}) &= \frac{1}{\delta_{Pjt}} \left(\delta_{Ojt} + \ln \frac{Q_{jt}^*}{\exp(\delta_{Mjt}) - Q_{jt}^*} + \frac{\exp(\delta_{Mjt})}{\exp(\delta_{Mjt}) - Q_{jt}^*}\right) \\ &= \frac{1}{\delta_{Pjt}} \left(\delta_{Ojt} + \ln \frac{S_{jt}^*}{1 - S_{jt}^*} + \frac{1}{1 - S_{jt}^*}\right), \end{aligned}$$

where $\delta_{jt} = (\delta_{Mjt}, \delta_{Ojt}, \delta_{Pjt})$ is the (three-dimensional) demand shock and $S_{jt}^* = \frac{Q_{jt}^*}{\exp(\delta_{Mjt})}$ is the market share of firm j based on planned output Q_{jt}^*. Consequently, if $\delta_{Ojt} = const$ and $\delta_{Pjt} = const$, then the market share S_{jt}^* controls for $MR(\cdot)$.

Different from CES demand, the markup $\mu_{jt} = \frac{\delta_{Pjt} P_{jt}(1-S^*_{jt})}{1+\delta_{Pjt} P_{jt}(1-S^*_{jt})}$ is not necessarily constant even if $\delta_{Ojt} = const$ and $\delta_{Pjt} = const$. In practice, however, it is unclear what allows the researcher to assume that market share S^*_{jt} is observed and measured without error when planned output Q^*_{jt} is unobserved and actual output $Q_{jt} = Q^*_{jt} \exp(\varepsilon_{jt})$ is measured with error. Calculating market share S^*_{jt} moreover requires the researcher to first define the market, which is not an easy undertaking, especially because production data typically covers only a sample of firms.

If $\delta_{Ojt} \neq const$ or $\delta_{Pjt} \neq const$, then the market share S^*_{jt} no longer suffices to control for $MR(\cdot)$. Note that $\delta_{Ojt} \neq const$ if rivals' prices are unobserved but cannot be assumed constant. Moreover, even if rivals' prices are observed, then $\delta_{Ojt} \neq const$ or $\delta_{Pjt} \neq const$ in a random-coefficients model along the lines of Berry et al. (1995).

Example 3: Cournot competition with homogeneous products. Market demand is $Q^*_t = D(P_t, \delta_t)$, where $Q^*_t = \sum_{j=1}^{N_t} Q^*_{jt}$ is the total planned output of the N_t firms and P_t the market price. Because all firms face the same demand, δ_t represents unobserved demand heterogeneity across time but not across firms. Inverse market demand is $P_t = P(Q^*_t, \delta_t)$ and the marginal revenue of firm j is

$$MR(Q^*_{jt}, \delta_t) = P(Q^*_t, \delta_t) + \frac{\partial P(Q^*_t, \delta_t)}{\partial Q^*_t} Q^*_{jt} = P_t \left(1 + \frac{S^*_{jt}}{\eta(P_t, \delta_t)}\right), \qquad (23)$$

where $S^*_{jt} = \frac{Q^*_{jt}}{Q^*_t}$ is the market share of firm j and $\eta(P_t, \delta_t) = \frac{\partial D(P_t,\delta_t)}{\partial P_t} \frac{P_t}{D(P_t,\delta_t)}$ the price elasticity of market demand. Consequently, time dummies and the market share S^*_{jt} control for $MR(\cdot)$. The same caveats as above apply regarding observing and calculating market share S^*_{jt}. Note also that Cournot competition with homogeneous products implies that the markup $\mu_{jt} = \frac{\eta(P_t,\delta_t)}{S^*_{jt}+\eta(P_t,\delta_t)}$ is decreasing in the market share S^*_{jt}. This may not be borne out by the estimated markup from the DLW method.

In sum, whether the endeavor of controlling for planned output Q^*_{jt} or, equivalently, for marginal cost $MC(\cdot)$ by observables z_{jt} is successful hinges on the specification of demand and assumptions on firm conduct. This reinforces our point that the DLW method does not free the researcher from having to think carefully about demand and firm conduct.

7 Dynamic panel method and alternative approaches

In light of the practical difficulties inherent in controlling for unobserved demand heterogeneity by observables z_{jt} and the assumptions required for this endeavor to succeed, a natural question is if there are alternatives to the proxy variable paradigm that are robust to unobserved demand heterogeneity. In this section, we illustrate the dynamic panel approach (Arellano & Bond 1991, Arellano & Bover 1995, Blundell & Bond 1998, Blundell & Bond 2000) and point to other alternative approaches to estimation.

We continue to assume an $AR(1)$ process with parameter ρ for Hicks-neutral productivity ω_{jt}. Taking logs and quasi-differentiating the production function in equation (1) yields the estimation equation

$$q_{jt} = \rho q_{jt-1} + \ln F(K_{jt}, L_{jt}, M_{jt}) - \rho \ln F(K_{jt-1}, L_{jt-1}, M_{jt-1}) + \xi_{jt} + \varepsilon_{jt} - \rho\varepsilon_{jt-1}. \tag{24}$$

Because the composite error term contains only the productivity innovation ξ_{jt} in addition to the current and lagged disturbances ε_{jt} and ε_{jt-1}, lagged inputs remain valid instruments, as does current capital K_{jt} under the assumption that the firm decides on investment, and thus capital, in period $t-1$ before it observes ω_{jt}.

Importantly, the dynamic panel approach avoids the inversion in the first step of ACF and therefore introducing δ_{jt} into the estimation.[25] Hence, the dynamic panel approach is robust to $\delta_{jt} \neq 0$ and offers a solution to the problem of unobserved demand heterogeneity and changes in firm conduct. Moreover, it offers a solution that does not require the researcher to control for any differences in demand and conduct across firms and time.

Ackerberg (2020) provides a detailed comparison of the OP/LP procedure and the dynamic panel approach. We highlight three disadvantages of the dynamic panel approach. First, it requires a stronger assumption on the stochastic process governing productivity than the OP/LP procedure. Second, because the dynamic panel approach avoids the inversion in the first step of ACF, it does not yield an estimate of the disturbance ε_{jt}. It is thus not able to estimate the markup μ_{jt} separately from the disturbance ε_{jt}.

[25] The dynamic panel approach may be viewed as rewriting the production function in equation (1) as $\omega_{jt} + \varepsilon_{jt} = q_{jt} - \ln F(K_{jt}, L_{jt}, L_{jt})$ and therefore as expressing $\omega_{jt} + \varepsilon_{jt}$ in terms of observables. The OP/LP procedure, in contrast, requires expressing ω_{jt} in terms of observables.

Third, in practice the dynamic panel approach may yield low or even negative estimates of the capital coefficient (De Loecker & Syverson 2021, p. 170).

Specification and estimation. Doraszelski & Jaumandreu (2019) discuss the importance of flexibly specifying the production function $F(K_{jt}, L_{jt}, M_{jt})$. We specify it to include a constant, a set of 21 year dummies, and a complete polynomial of order 3 in k_{jt}, l_{jt}, and m_{jt} to allow output elasticities to vary with inputs. We estimate equation (24) by GMM. In addition to the constant and the year dummies, the instruments are a complete polynomial of order 3 in k_{jt-1}, l_{jt-1}, and m_{jt-1} and a polynomial of order 3 in k_{jt}. We winsorize the estimates of the output elasticity of capital $\beta_K(K_{jt}, L_{jt}, M_{jt})$ and the short-run elasticity of scale $\nu(K_{jt}, L_{jt}, M_{jt})$ at the 0.05 and 0.95 percentiles.[26]

Markup. Because the dynamic panel approach does not yield an estimate of the disturbance ε_{jt}, we rewrite equation (21) as

$$\ln \mu_{jt} + \varepsilon_{jt} = \ln \nu(K_{jt}, L_{jt}, M_{jt}) - \ln \left(S^R_{Ljt} + S^R_{Mjt}\right). \tag{25}$$

We estimate $\ln \mu_{jt} + \varepsilon_{jt}$ by substituting the estimate of the short-run elasticity of scale $\ln \nu(K_{jt}, L_{jt}, M_{jt})$ into equation (25). Note that any average of $\ln \mu_{jt} + \varepsilon_{jt}$ over a sufficiently large number of firms and/or years is a consistent estimate of the average (log) markup for these firms and/or years.

Results. Table 3 reports the results from the dynamic panel approach. Column (1) shows the average (log) markup by industry, along with the sample standard deviation. Due to the flexible specification of the production function, the sample standard deviation is large. Yet, with the exception of industry 9, the average markup is sensible. It is noticeably smaller in industries 3, 4, and 5 compared to the average markup obtained by the DLW method in column (1) of Table 1.

Columns (2)–(4) of Table 3 show the average estimate of the output elasticity of capital $\beta_K(\cdot)$, the average estimate of the short-run elasticity of scale $\nu(\cdot)$, and the estimate of the $AR(1)$ parameter ρ. The average estimate

[26] This eliminates some extreme values without materially affecting the mean of the distribution.

Table 3: Dynamic panel method

	Baseline specification				Sargan test
	Markup C (s. dev.)	$\beta(\cdot)$	$\nu(\cdot)$	ρ (s. e.)	p-val. (d. f.)
	(1)	(2)	(3)	(4)	(5)
1. Metals and metal products	0.116	0.047	1.036	0.820	0.896
	(0.057)			(0.086)	(2)
2. Non-metallic minerals	0.235	0.022	1.024	0.551	0.838
	(0.076)			(0.071)	(2)
3. Chemical products	0.213	0.053	0.955	0.781	0.547
	(0.026)			(0.652)	(2)
4. Agric. and ind. machinery	0.214	0.003	0.993	0.787	0.281
	(0.045)			(0.039)	(5)
5. Electrical goods	0.214	0.004	0.995	0.785	0.537
	(0.038)			(0.037)	(8)
6. Transport equipment	0.161	0.064	0.975	0.262	0.779
	(0.076)			(0.142)	(2)
7. Food, drink and tobacco	0.228	0.065	0.995	0.779	0.981
	(0.041)			(0.144)	(2)
8. Textile, leather and shoes	0.229	-0.001	1.063	0.725	0.338
	(0.089)			(0.036)	(5)
9. Timber and furniture	0.041	0.005	1.006	0.778	0.911
	(0.082)			(0.038)	(5)
10. Paper and printing products	0.224	0.086	1.054	0.788	0.832
	(0.024)			(0.023)	(5)

of the output elasticity of capital $\beta_K(\cdot)$ is on the low side, especially in industries 4, 5, 8, and 9. The average estimate of the short-run elasticity of scale $\nu(\cdot)$ is on the high side and ranges from 0.955 to 1.100, comparable to column (2) of Table 1.[27] The Sargan test in column (5) does not reject the specification in any industry at conventional significance levels.

Alternative approaches. While the dynamic panel method is well-established for estimating production functions, recent work aims to alleviate some of its drawbacks, notably the stronger assumption on the stochastic process governing productivity. Ponder (2021) applies the work on polynomial errors-in-variables models by Hausman, Newey, Ichimura & Powell (1991) to show that the dynamic panel approach can be extended from an $AR(1)$ process for Hicks-neutral productivity ω_{jt} to a Markov process with a polynomial law of motion.

Abito (2022) similarly assumes a Markov process with a polynomial law of motion and draws on Hausman et al. (1991) to show how to incorporate firm fixed effects into the proxy variable paradigm. If differences in demand across time within firms can be ruled out and if demand satisfies a separability assumption, then these fixed effects control for the remaining unobserved demand heterogeneity.

Brand (2020) applies the work on nonclassical measurement error by Hu & Schennach (2008) to propose an alternative identification and estimation procedure that avoids the inversion in the first step of ACF and instead treats the firm's output conditional on its inputs as a noisy signal of its productivity.

Demirer (2019) develops a partial identification approach that treats either investment or materials as an "imperfect proxy" for productivity that can additionally depend on a demand shock. His setup gives rise to moment inequalities if the productivity distribution conditional on the imperfect proxy being above a threshold stochastically dominates the productivity distribution conditional on the imperfect proxy being below the threshold. Of course, these alternative approaches come with their own drawbacks and more work is required to assess how well they work in practice.

[27] We note that if we instead specify a Cobb-Douglas production function, then in 9 industries we estimate the output elasticity of capital to be even lower than the one we report in Table 3 while we estimate the short-run elasticity of scale to be comparable.

8 Concluding remarks

DLW obtain the markup from the firm's cost minimization problem by substituting in estimates of the output elasticity of a variable input and the disturbance that separates actual from planned output. These estimates are obtained using the OP/LP procedure. Our paper has highlighted the under-appreciated assumption of the DLW method that to consistently estimate markups, it either has to rule out any differences in demand and conduct across firms and time or assume that they can be fully controlled for by observables z_{jt}.

The demand a firm faces depends not only on the product characteristics of the firm but, in imperfectly competitive industries, also on the product characteristics of its rivals and their prices or quantities. Because typical production data has even less information on demand than our market dynamism variable and rivals are partially or completely unobserved, attempting to control for any differences in demand across firms and time by observables z_{jt} is difficult in practice. It is also unappealing from a conceptual point of view. The demand a firm faces and its conduct are the fundamental determinants of the markup it charges. Hence, to use an OP/LP procedure to estimate the production function and obtain the markup, the DLW method would have to observe and control for all these determinants of the markup.

To develop the consequences of unobserved demand heterogeneity and changes in firm conduct, we have characterized the bias in the estimates produced by the DLW method in the presence of a demand shock $\delta_{jt} \neq 0$. The bias permeates the level of the estimated markup and its correlation with variables of interest. We have shown that both the level and the correlation component of the bias can be severe. We have provided an empirical application to test for the effects of unobserved demand heterogeneity. In our application, the bias is most pronounced in the correlation of the estimated markup with our market dynamism variable. Similar correlations of the estimated markup with variables of interest are often the focus of attention in applications of DLW.

We have further used our empirical application to illustrate the dynamic panel approach to estimation as an alternative to the proxy variable paradigm. Because the dynamic panel approach avoids the inversion in the first step of ACF and therefore introducing δ_{jt} into the estimation, it offers a solution to the problem of unobserved demand heterogeneity and changes in firm conduct. Moreover, it offers a solution that does not require the researcher to control for any differences in demand and conduct across firms and time. The disadvantages of the dynamic panel approach include the requirement to assume an $AR(1)$ process for Hicks-neutral productivity ω_{jt} and the inability to estimate the markup μ_{jt} separately from the disturbance ε_{jt}. Despite these disadvantages, there is little reason not to use the dynamic panel approach at least to alleviate concerns about unobserved demand heterogeneity and changes in firm conduct.

In addition to unobserved demand heterogeneity and changes in firm conduct, the production approach to estimating the markup has other issues to confront, including flexibly specifying the production function (Doraszelski & Jaumandreu 2019, Demirer 2020), the costly adjustment of inputs, and market power in input markets (Yeh, Macaluso & Hershbein 2022, Rubens 2023, Azzam, Jaumandreu & Lopez 2023). A particularly important consideration is that consistently estimating the output elasticity may be difficult in a model that restricts productivity to be single-dimensional. A number of recent papers provide evidence of labor-augmenting productivity (Doraszelski & Jaumandreu 2018, Raval 2019, Zhang 2019, Demirer 2020). In contrast to Hicks-neutral productivity, labor-augmenting productivity directly enters the output elasticity. The literature has only recently begun to develop more sophisticated models and estimators to handle multi-dimensional productivity. Doraszelski & Jaumandreu (2019), Demirer (2020), and Raval (2022) in particular highlight the implications of biased technological change for markup estimation.

In sum, contrary to the purported advantage of relying on cost minimization, the DLW method does not free the researcher from having to think carefully about the specification of demand and assumptions on firm conduct. The conditions required by the DLW method to consistently estimate markups may be difficult to satisfy. As we have shown, violations of these conditions can be consequential.

Appendix A

We consider a merger between two symmetric Bertrand competitors labelled j and $-j$. The production function of firm j is $Q^*_{jt} = V_{jt}^{\beta_V} \exp(\omega_{jt})$. The cost of producing planned output Q^*_{jt} is therefore $C_{jt} = \left(Q^*_{jt}\right)^{\frac{1}{\beta_V}} \exp\left(-\frac{\omega_{jt}}{\beta_V}\right)$, where we use $P_{Vjt} = 1$. The demand firm j faces is $Q^*_{jt} = \exp(\delta_{jt}) P^{\eta}_{jt} P^{\gamma}_{-jt}$. The production, cost, and demand functions of firm $-j$ are analogous.

The profit of firm j is $\pi_{jt} = P_{jt} Q^*_{jt} - C_{jt}$. We model a merger between firm j and firm $-j$ by assuming that firm j assigns weight $\iota_{jt} \in \{0, 1\}$ to the profit of firm $-j$ and maximizes $\pi_{jt} + \iota_{jt}\pi_{-jt}$. Pre-merger $\iota_{jt} = 0$ and firm j maximizes π_{jt}; post-merger $\iota_{jt} = 1$ and firm j maximizes $\pi_{jt} + \pi_{-jt}$.

The FOC for the price of output P_{jt} is

$$(\eta+1)\exp(\delta_{jt})P^{\eta}_{jt}P^{\gamma}_{-jt} - \frac{1}{\beta_V}\left(\exp(\delta_{jt})P^{\eta}_{jt}P^{\gamma}_{-jt}\right)^{\frac{1-\beta_V}{\beta_V}} \exp\left(-\frac{\omega_{jt}}{\beta_V}\right)\eta \exp(\delta_{jt})P^{\eta-1}_{jt}P^{\gamma}_{-jt}$$

$$+\iota_{jt}\left(\gamma\exp(\delta_{-jt})P^{\eta+1}_{-jt}P^{\gamma-1}_{jt} - \frac{1}{\beta_V}\left(\exp(\delta_{-jt})P^{\eta}_{-jt}P^{\gamma}_{jt}\right)^{\frac{1-\beta_V}{\beta_V}} \exp\left(-\frac{\omega_{-jt}}{\beta_V}\right)\gamma\exp(\delta_{-jt})P^{\eta}_{-jt}P^{\gamma-1}_{jt}\right) = 0.$$

In the interest of tractability, we assume $\omega_{jt} = \omega_{-jt}$ and $\delta_{jt} = \delta_{-jt}$ and hence $P_{jt} = P_{-jt}$ in equilibrium.[28] The FOC accordingly simplifies to

$$\frac{\eta+\iota_{jt}\gamma+1}{\eta+\iota_{jt}\gamma}P_{jt} = \frac{1}{\beta_V}\left(\exp(\delta_{jt})P^{\eta+\gamma}_{jt}\right)^{\frac{1-\beta_V}{\beta_V}}\exp\left(-\frac{\omega_{jt}}{\beta_V}\right). \tag{26}$$

Because the right-hand side of equation (26) is marginal cost MC_{jt} at planned output $Q^*_{jt} = \exp(\delta_{jt})P^{\eta+\gamma}_{jt}$, in equilibrium the markup is $\mu_{jt} = \frac{\eta+\iota_{jt}\gamma}{\eta+\iota_{jt}\gamma+1}$.

Appendix B

Data generating process. Solving equation (26) yields

$$p_{jt} = \frac{\beta_V}{(1-\beta_V)(\eta+\gamma)-\beta_V}\left(\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma} - \frac{(1-\beta_V)\delta_{jt}}{\beta_V} + \frac{\omega_{jt}}{\beta_V}\right) \tag{27}$$

and thus

$$q^*_{jt} = \frac{\beta_V(\eta+\gamma)}{(1-\beta_V)(\eta+\gamma)-\beta_V}\left(\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma} - \frac{\delta_{jt}}{\eta+\gamma} + \frac{\omega_{jt}}{\beta_V}\right), \tag{28}$$

[28] We may assume that firm j is in our sample whereas firm $-j$ is not, as is typical in production data.

$$v_{jt} = \frac{\eta+\gamma}{(1-\beta_V)(\eta+\gamma)-\beta_V}\left(\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma} - \frac{\delta_{jt}}{\eta+\gamma} + \frac{(\eta+\gamma+1)\omega_{jt}}{\eta+\gamma}\right). \tag{29}$$

Equations (27)–(29) generate the endogenous variables q^*_{jt}, v_{jt}, and p_{jt} from the exogenous variables ω_{jt}, δ_{jt}, and ι_{jt}.

Gaussian mixture model. Because conditional on ι_{jt} the expressions for q^*_{jt}, v_{jt}, and p_{jt} in equations (27)–(29) are linear in ω_{jt} and δ_{jt}, the assumption that $\omega_{jt} \sim N(0, \sigma^2_\omega)$ and $\delta_{jt} \sim N(0, \sigma^2_\delta)$ gives rise to a Gaussian mixture model. Conditional on ι_{jt}, we have

$$\begin{pmatrix} q^*_{jt} \\ v_{jt} \\ p_{jt} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{q^*|\iota_{jt}} \\ \mu_{v|\iota_{jt}} \\ \mu_{p|\iota_{jt}} \end{pmatrix}, \begin{pmatrix} \Sigma_{q^*,q^*} & \Sigma_{q^*,v} & \Sigma_{q^*,p} \\ \Sigma_{q^*,v} & \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{q^*,p} & \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}\right) = N\left(\mu_{\iota_{jt}}, \Sigma\right),$$

where

$$\mu_{q^*|\iota_{jt}} = \frac{\beta_V(\eta+\gamma)}{(1-\beta_V)(\eta+\gamma)-\beta_V}\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma}, \tag{30}$$

$$\mu_{v|\iota_{jt}} = \frac{\eta+\gamma}{(1-\beta_V)(\eta+\gamma)-\beta_V}\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma}, \tag{31}$$

$$\mu_{p|\iota_{jt}} = \frac{\beta_V}{(1-\beta_V)(\eta+\gamma)-\beta_V}\ln\frac{\beta_V(\eta+\iota_{jt}\gamma+1)}{\eta+\iota_{jt}\gamma}, \tag{32}$$

$$\Sigma_{q^*,q^*} = \left(\frac{\beta_V(\eta+\gamma)}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{\sigma^2_\delta}{(\eta+\gamma)^2} + \frac{\sigma^2_\omega}{\beta^2_V}\right),$$

$$\Sigma_{q^*,v} = \beta_V\left(\frac{\eta+\gamma}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{\sigma^2_\delta}{(\eta+\gamma)^2} + \frac{(\eta+\gamma+1)\sigma^2_\omega}{\beta_V(\eta+\gamma)}\right),$$

$$\Sigma_{q^*,p} = (\eta+\gamma)\left(\frac{\beta_V}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{(1-\beta_V)\sigma^2_\delta}{\beta_V(\eta+\gamma)} + \frac{\sigma^2_\omega}{\beta^2_V}\right),$$

$$\Sigma_{v,v} = \left(\frac{\eta+\gamma}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{\sigma^2_\delta}{(\eta+\gamma)^2} + \frac{(\eta+\gamma+1)^2\sigma^2_\omega}{(\eta+\gamma)^2}\right),$$

$$\Sigma_{v,p} = \beta_V(\eta+\gamma)\left(\frac{1}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{(1-\beta_V)\sigma^2_\delta}{\beta_V(\eta+\gamma)} + \frac{(\eta+\gamma+1)\sigma^2_\omega}{\beta_V(\eta+\gamma)}\right),$$

$$\Sigma_{p,p} = \left(\frac{\beta_V}{(1-\beta_V)(\eta+\gamma)-\beta_V}\right)^2\left(\frac{(1-\beta_V)^2\sigma^2_\delta}{\beta^2_V} + \frac{\sigma^2_\omega}{\beta^2_V}\right).$$

Unconditionally, we have a multivariate Gaussian mixture with density

$$(1-\tau)\mathcal{N}\left(\begin{pmatrix} q^*_{jt} \\ v_{jt} \\ p_{jt} \end{pmatrix}; \mu_{\iota_{jt}=0}, \Sigma\right) + \tau\mathcal{N}\left(\begin{pmatrix} q^*_{jt} \\ v_{jt} \\ p_{jt} \end{pmatrix}; \mu_{\iota_{jt}=1}, \Sigma\right),$$

where $\lambda = \Pr(\iota_{jt} = 1) \in (0,1)$ and $\mathcal{N}(\cdot; \mu_{\iota_{jt}}, \Sigma)$ is a multivariate normal density with parameters $\mu_{\iota_{jt}}$ and Σ.

First step of ACF. The first step of ACF estimates the conditional expectation $E(q_{jt}|z_{jt}) = E(q^*_{jt}|z_{jt})$, where $z_{jt} = (v_{jt}, p_{jt})$. Normality implies that, conditional on (z_{jt}, ι_{jt}),[29]

$$\begin{aligned} E(q^*_{jt}|z_{jt}, \iota_{jt}) &= \mu_{q^*|\iota_{jt}} + \begin{pmatrix} \Sigma_{q^*,v} & \Sigma_{q^*,p} \end{pmatrix} \begin{pmatrix} \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}^{-1} \begin{pmatrix} v_{jt} - \mu_{v|\iota_{jt}} \\ p_{jt} - \mu_{p|\iota_{jt}} \end{pmatrix} \\ &= -\ln \frac{\beta_V(\eta + \iota_{jt}\gamma + 1)}{\eta + \iota_{jt}\gamma} + v_{jt} - p_{jt}. \end{aligned} \tag{33}$$

This implies that, conditional on z_{jt},

$$\begin{aligned} E(q^*_{jt}|z_{jt}) = E&\left(-\ln \frac{\beta_V(\eta + \iota_{jt}\gamma + 1)}{\eta + \iota_{jt}\gamma} + v_{jt} - p_{jt}\middle| z_{jt}, \iota_{jt} = 0\right) \Pr(\iota_{jt} = 0|z_{jt}) \\ +E&\left(-\ln \frac{\beta_V(\eta + \iota_{jt}\gamma + 1)}{\eta + \iota_{jt}\gamma} + v_{jt} - p_{jt}\middle| z_{jt}, \iota_{jt} = 1\right) \Pr(\iota_{jt} = 1|z_{jt}) \\ &= -\ln \beta_V + v_{jt} - p_{jt} + \ln \underline{\mu} + \left(\ln \overline{\mu} - \ln \underline{\mu}\right) \tau(z_{jt}), \end{aligned} \tag{34}$$

where $\underline{\mu} = \frac{\eta}{\eta+1}$, $\overline{\mu} = \frac{\eta+\gamma}{\eta+\gamma+1}$, and

$$\begin{aligned} \tau(z_{jt}) = \Pr(\iota_{jt} = 1|z_{jt}) = \lambda\mathcal{N}\left(z_{jt}; \begin{pmatrix} \mu_{v|\iota_{jt}=1} \\ \mu_{p|\iota_{jt}=1} \end{pmatrix}, \begin{pmatrix} \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}\right) \\ \Big/ \left((1-\lambda)\mathcal{N}\left(z_{jt}; \begin{pmatrix} \mu_{v|\iota_{jt}=0} \\ \mu_{p|\iota_{jt}=0} \end{pmatrix}, \begin{pmatrix} \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}\right) + \lambda\mathcal{N}\left(z_{jt}; \begin{pmatrix} \mu_{v|\iota_{jt}=1} \\ \mu_{p|\iota_{jt}=1} \end{pmatrix}, \begin{pmatrix} \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}\right)\right) \\ \in (0,1). \end{aligned} \tag{35}$$

The first equality in equation (35) defines $\tau(z_{jt})$ as shorthand for $\Pr(\iota_{jt} = 1|z_{jt})$ and the second equality uses Bayes' theorem to provide an expression for $\Pr(\iota_{jt} = 1|z_{jt})$. Because $\tau(z_{jt}) \in (0,1)$, we think of it as a weight.

[29] Equation (33) can alternatively be derived without invoking normality by using equations (27) and (29) to eliminate ω_{jt} and δ_{jt} in equation (28).

Markup regression. Regressing the markup $\ln \mu_{jt}^{DLW} = \ln \mu_{jt} - \zeta_{jt}$ obtained by the DLW method on a constant and the step dummy ι_{jt} estimates

$$E\left(\ln \mu_{jt}^{DLW} | \iota_{jt}\right) = E\left(\ln \mu_{jt}^{DLW} | \iota_{jt} = 0\right) + \left(E\left(\ln \mu_{jt}^{DLW} | \iota_{jt} = 1\right) - E\left(\ln \mu_{jt}^{DLW} | \iota_{jt} = 0\right)\right) \iota_{jt},$$

where

$$E\left(\ln \mu_{jt}^{DLW} | \iota_{jt}\right) = E\left(\ln \mu_{jt} | \iota_{jt}\right) - E\left(\zeta_{jt} | \iota_{jt}\right) = \ln \frac{\eta + \iota_{jt}\gamma}{\eta + \iota_{jt}\gamma + 1} - E(\zeta_{jt} | \iota_{jt}). \tag{36}$$

Turning to the prediction error $\zeta_{jt} = q_{jt}^* - E(q_{jt}^* | z_{jt})$, using equation (34), $E(q_{jt}^* | \iota_{jt}) = \mu_{q^* | \iota_{jt}}$, $E(v_{jt} | \iota_{jt}) = \mu_{v | \iota_{jt}}$, $E(p_{jt} | \iota_{jt}) = \mu_{p | \iota_{jt}}$, and the expressions for $\mu_{q^* | \iota_{jt}}$, $\mu_{v | \iota_{jt}}$, and $\mu_{p | \iota_{jt}}$ in equations (30)–(32) yields

$$\begin{aligned} E\left(\zeta_{jt} | \iota_{jt}\right) &= E\left(q_{jt}^* | \iota_{jt}\right) - E\left(E(q_{jt}^* | z_{jt}) | \iota_{jt}\right) \\ &= E\left(q_{jt}^* | \iota_{jt}\right) - E\left(-\ln \beta_V + v_{jt} - p_{jt} + \ln \underline{\mu} + \left(\ln \overline{\mu} - \ln \underline{\mu}\right) \tau(z_{jt}) \middle| \iota_{jt}\right) \\ &= \mu_{q^* | \iota_{jt}} + \ln \beta_V - \mu_{v | \iota_{jt}} + \mu_{p | \iota_{jt}} - \ln \underline{\mu} - \left(\ln \overline{\mu} - \ln \underline{\mu}\right) E\left(\tau(z_{jt}) | \iota_{jt}\right) \\ &= \ln \frac{\eta + \iota_{jt}\gamma}{\eta + \iota_{jt}\gamma + 1} - \ln \underline{\mu} - \left(\ln \overline{\mu} - \ln \underline{\mu}\right) E(\tau(z_{jt}) | \iota_{jt}), \end{aligned} \tag{37}$$

where

$$E(\tau(z_{jt}) | \iota_{jt}) = \int \tau(z_{jt}) \mathcal{N}\left(z_{jt}; \begin{pmatrix} \mu_{v | \iota_{jt}} \\ \mu_{p | \iota_{jt}} \end{pmatrix}, \begin{pmatrix} \Sigma_{v,v} & \Sigma_{v,p} \\ \Sigma_{v,p} & \Sigma_{p,p} \end{pmatrix}\right) dz_{jt}. \tag{38}$$

Turning back to the markup $\ln \mu_{jt}^{DLW} = \ln \mu_{jt} - \zeta_{jt}$ obtained by the DLW method, substituting equation (37) into equation (36) yields

$$E\left(\ln \mu_{jt}^{DLW} | \iota_{jt}\right) = \ln \underline{\mu} + \left(\ln \overline{\mu} - \ln \underline{\mu}\right) E(\tau(z_{jt}) | \iota_{jt}).$$

The coefficient of interest is thus

$$\begin{aligned} &E\left(\ln \mu_{jt}^{DLW} | \iota_{jt} = 1\right) - E\left(\ln \mu_{jt}^{DLW} | \iota_{jt} = 0\right) \\ &= \left(\ln \overline{\mu} - \ln \underline{\mu}\right) \left(E(\tau(z_{jt}) | \iota_{jt} = 1) - E(\tau(z_{jt}) | \iota_{jt} = 0)\right). \end{aligned} \tag{39}$$

Equation (35) implies $\tau(z_{jt}) = 0$ if $\lambda = 0$ and $\tau(z_{jt}) = 1$ if $\lambda = 1$. The coefficient of interest in equation (39) is therefore estimated to be zero in these extreme cases. To evaluate the bias for intermediate cases, we parameterize

$$\beta_V = 0.8, \quad \eta = -6, \quad \gamma = 2.5, \quad \sigma_\omega^2 = 0.25, \quad \sigma_\delta^2 \in \{0.01, 0.05, 0.25, 1.25\}.$$

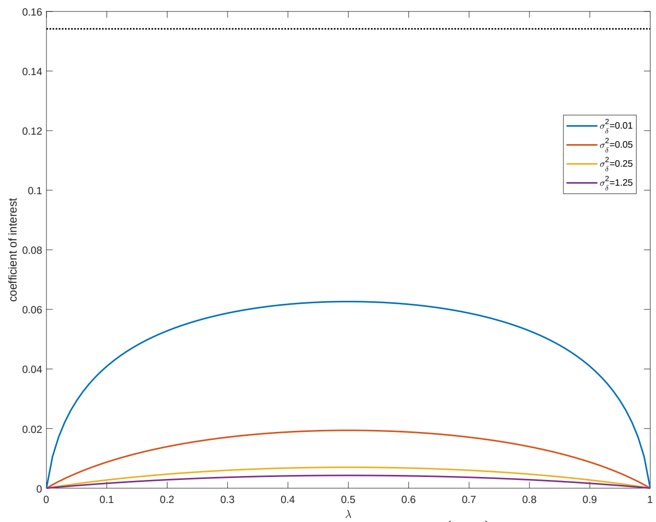

Figure 2: Coefficient of interest in equation (39) over range of λ for $\sigma_\delta^2 \in \{0.01, 0.05, 0.25, 1.25\}$.

The wave of acquisitions may therefore enable a firm to increase its markup from $\underline{\mu} = 1.2$ to $\overline{\mu} = 1.4$. Total factor productivity $\exp(\omega_{jt})$ is 3.60 times larger at the 90th percentile of the productivity distribution than at the 10th percentile.[30] The demand shock δ_{jt} may be less, equally, or more dispersed than Hicks-neutral productivity ω_{jt}. Figure 2 shows the coefficient of interest over the entire range of λ for the various values of σ_δ^2. By comparison, regressing $\ln \mu_{jt}$ instead of $\ln \mu_{jt}^{DLW}$ on a constant and the step dummy ι_{jt} yields $\ln 1.4 - \ln 1.2 = 0.1542$ for the coefficient of interest. As can be seen, even for very small values of σ_δ^2, the coefficient of interest is severely biased towards zero over the entire range of λ.

Excluding price of output. Excluding the price of output P_{jt} from the first step of ACF, we specify $z_{jt} = v_{jt}$. Normality implies that, conditional on (z_{jt}, ι_{jt}),

$$E(q_{jt}^*|z_{jt}, \iota_{jt}) = \mu_{q^*|\iota_{jt}} + \Sigma_{q^*,v}\Sigma_{v,v}^{-1}\left(z_{jt} - \mu_{v|\iota_{jt}}\right)$$

[30] Syverson (2004) cites 90:10 ratios of about 2 and Hsieh & Klenow (2009) of about 5.

$$= \frac{(\eta+\gamma+1)(\eta+\gamma)\sigma_\omega^2\left(-\ln\frac{\beta_V(1+\eta+\iota_{jt}\gamma)}{\eta+\iota_{jt}\gamma}+v_{jt}\right)+\sigma_\delta^2\beta_V v_{jt}}{(\eta+\gamma+1)^2\sigma_\omega^2+\sigma_\delta^2}.$$

The derivations proceed along similar lines as before. We obtain $E\left(\ln\mu_{jt}^{DLW}|\iota_{jt}=1\right)-E\left(\ln\mu_{jt}^{DLW}|\iota_{jt}=0\right)=-0.0508$ if we parameterize $\sigma_\delta^2=0.01$ and $\lambda=0.5$. Hence, the coefficient of interest may be estimated to be negative instead of positive.

Appendix C

To simplify the exposition, we omit firm and time subscripts in what follows. We use the superscript $'$ to denote a lead, the subscript -1 to denote a first lag, and the subscript -2 to denote a second lag, etc.

Data generating process: variable input. Given capital K, the demand for the variable input is $V=\left(\frac{Q^*}{K^{\beta_K}\exp(\omega)}\right)^{\frac{1}{\beta_V}}$. Using equation (17) thus yields

$$v=\frac{1}{\beta_V}\left(\frac{1-\beta_V}{\beta_V}(\beta_K k+\omega)-p_V+\delta\right). \tag{40}$$

Variable cost is $VC=P_V\left(\frac{Q^*}{K^{\beta_K}\exp(\omega)}\right)^{\frac{1}{\beta_V}}$. Marginal cost is $MC=\frac{1}{\beta_V}P_V\frac{(Q^*)^{\frac{1-\beta_V}{\beta_V}}}{\left(K^{\beta_K}\exp(\omega)\right)^{\frac{1}{\beta_V}}}$ and thus depends on the index $\frac{1}{\beta_V}(\beta_K k-\beta_V p_V+\omega)$.

Data generating process: capital. The firm's dynamic programming problem simplifies considerably if capital is subject to time to build but not to adjustment costs (Adda & Cooper 2003, pp. 189–190). Capital evolves as $K'=(1-\theta)K+I$, where $\theta\in[0,1]$ is the rate of depreciation and I is investment. The firm buys and sells capital in the spot market each period.

In our setup, in the current period the firm sets K' to minimize its expected total cost in the subsequent period while accounting for the resale value of undepreciated capital:

$$\min_{K'} E\left(VC'+P_K K'-P_K'(1-\theta)K'|Q^*,K',K,P_K,P_V,\omega,\delta\right),$$

where $P_K K'$ is the acquisition cost of capital and $P_K'(1-\theta)K'$ is the resale value of undepreciated capital. The FOC can be written as[31]

$$\ln \frac{\beta_K}{\beta_V} + \ln E\left(P_V'\left(\frac{Q^{*\prime}}{\exp(\omega')}\right)^{\frac{1}{\beta_V}} + P_K'(1-\theta)|Q^*, K', K, P_K, P_V, \omega, \delta\right) - \frac{\beta_K+\beta_V}{\beta_V}k' = p_K. \tag{41}$$

Setting $\theta = 1$ for simplicity, we have

$$\begin{aligned} & E\left(P_V'\left(\frac{Q^{*\prime}}{\exp(\omega')}\right)^{\frac{1}{\beta_V}} |Q^*, K', K, P_K, P_V, \omega, \delta\right) \\ = & E\left(\exp\left(p_V' + \frac{1}{\beta_V}\left(\frac{1}{\beta_V}\left(\beta_K k' - \beta_V p_V' + \omega'\right) + \delta' - \omega'\right)\right)|Q^*, K', K, P_K, P_V, \omega, \delta\right) \\ = & \exp\left(\frac{\beta_K}{\beta_V^2}k' - \frac{1-\beta_V}{\beta_V}\rho_{p_V} p_V + \frac{1-\beta_V}{\beta_V^2}\rho_\omega \omega + \frac{1}{\beta_V}\rho_\delta \delta\right) E\left(e^{-\frac{1-\beta_V}{\beta_V}\xi_{p_V}' + \frac{1-\beta_V}{\beta_V^2}\xi_\omega' + \frac{1}{\beta_V}\xi_\delta'}\right), \end{aligned}$$

where the first equality uses equation (17) and the second equality exploits the assumed $AR(1)$ processes for p_V, ω, and δ. Hence, we obtain

$$\begin{aligned} k' = & \frac{1}{\beta_V(\beta_K+\beta_V) - \beta_K} \\ & \cdot\left(\beta_V\left(\ln\frac{\beta_K}{\beta_V} + \ln E(\cdot)\right) - \beta_V p_K - (1-\beta_V)\rho_{p_V} p_V + \frac{1-\beta_V}{\beta_V}\rho_\omega \omega + \rho_\delta \delta\right), \end{aligned} \tag{42}$$

where, again exploiting the assumed $AR(1)$ processes,

$$\ln E(\cdot) = \left(\frac{1-\beta_V}{\beta_V}\right)^2 \frac{\sigma_{p_V}^2}{2} + \left(\frac{1-\beta_V}{\beta_V^2}\right)^2 \frac{\sigma_\omega^2}{2} + \left(\frac{1}{\beta_V}\right)^2 \frac{\sigma_\delta^2}{2}.$$

Equation (42) is a special case of the law of motion in equation (18). From hereon we work with the more general equation (18).

$VAR(1)$ **process.** Equations (17) and (40) generate q^* and v from k, p_V, ω, and δ, and equation (18) generates k' from k, p_K, p_V, ω, and δ. Combining

[31] As it takes planned output $Q^{*\prime}$ as given in the dynamic cost minimization problem, the firm presumes $\frac{\partial Q^{*\prime}}{\partial K'} = 0$.

these equations and the assumed $AR(1)$ processes for p_K, p_V, ω, and δ yields the $VAR(1)$ process

$$\begin{pmatrix} q^* \\ k' \\ v \\ p'_K \\ p'_V \\ \omega' \\ \delta' \end{pmatrix} = \begin{pmatrix} 0 \\ \tau_0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 & \frac{\beta_K}{\beta_V} & 0 & 0 & -1 & \frac{1}{\beta_V} & 1 \\ 0 & \tau_k & 0 & \tau_{p_K} & \tau_{p_V} & \tau_\omega & \tau_\delta \\ 0 & \frac{(1-\beta_V)\beta_K}{\beta_V^2} & 0 & 0 & -\frac{1}{\beta_V} & \frac{1-\beta_V}{\beta_V^2} & \frac{1}{\beta_V} \\ 0 & 0 & 0 & \rho_{p_K} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \rho_{p_V} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \rho_\omega & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_\delta \end{pmatrix} \begin{pmatrix} q^*_{-1} \\ k \\ v_{-1} \\ p_K \\ p_V \\ \omega \\ \delta \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ \xi'_{p_K} \\ \xi'_{p_V} \\ \xi'_\omega \\ \xi'_\delta \end{pmatrix}.$$

Defining $x = (q^*, k', v, p'_K, p'_V, \omega', \delta')^T$, $u = \left(0, 0, 0, \xi'_{p_K}, \xi'_{p_V}, \xi'_\omega, \xi'_\delta\right)^T$, and the coefficient vector a, respectively, matrix A accordingly, we write the $VAR(1)$ model more compactly as

$$x = a + Ax_{-1} + u, \quad u \sim N(0, \Upsilon),$$

where $\Upsilon = diag\left(0, 0, 0, \sigma^2_{p_K}, \sigma^2_{p_V}, \sigma^2_\omega, \sigma^2_\delta\right)$.

In what follows we exploit that our setup yields a $VAR(1)$ process to derive the various terms in equation (16)—and hence the bias in the estimate of the output elasticity of the variable input V—as functions of the primitives. This allows us to evaluate the bias and to prove Proposition 1–3 by direct calculation using a computer algebra system.

Expectations, covariances, and Yule-Walker equations. Define the expectation vector $\mu = E(x)$. Assuming stability of the $VAR(1)$ process, we obtain

$$\mu = (I - A)^{-1} a \tag{43}$$

in closed form (Lütkepohl 2005, pp. 15). We use μ_{q^*} as shorthand for $E(q^*)$, $\mu_{k'}$ for $E(k')$, etc.

Define the covariance matrix $\Gamma^h = Cov(x, x_{-h})$ for all $h \geq 0$. The Yule-Walker equations are

$$\Gamma^0 = A\Gamma^0 A^T + \Upsilon, \quad \Gamma^h = A\Gamma^{h-1}, \quad h \geq 1.$$

Solving these equations yields Γ^h for all $h \geq 0$ in closed form (Lütkepohl 2005, pp. 26–27). Given the timing of the various variables in $x = (q^*, k', v, p'_K, p'_V, \omega', \delta')^T$, we obtain

$$\Gamma^0 = \begin{pmatrix} \Sigma_{q^*,q^*} & \Sigma_{q^*,k'} & \Sigma_{q^*,v} & \Sigma_{q^*,p'_K} & \Sigma_{q^*,p'_V} & \Sigma_{q^*,\omega'} & \Sigma_{q^*,\delta'} \\ \dots & \Sigma_{k',k'} & \Sigma_{k',v} & \Sigma_{k',p'_K} & \Sigma_{k',p'_V} & \Sigma_{k',\omega'} & \Sigma_{k',\delta'} \\ \dots & \dots & \Sigma_{v,v} & \Sigma_{v,p'_K} & \Sigma_{v,p'_V} & \Sigma_{v,\omega'} & \Sigma_{v,\delta'} \\ \dots & \dots & \dots & \Sigma_{p'_K,p'_K} & \Sigma_{p'_K,p'_V} & \Sigma_{p'_K,\omega'} & \Sigma_{p'_K,\delta'} \\ \dots & \dots & \dots & \dots & \Sigma_{p'_V,p'_V} & \Sigma_{p'_V,\omega'} & \Sigma_{p'_V,\delta'} \\ \dots & \dots & \dots & \dots & \dots & \Sigma_{\omega',\omega'} & \Sigma_{\omega',\delta'} \\ \dots & \dots & \dots & \dots & \dots & \dots & \Sigma_{\delta',\delta'} \end{pmatrix}, \qquad (44)$$

$$\Gamma^1 = \begin{pmatrix} \Sigma_{q^*,q^*_{-1}} & \Sigma_{q^*,k} & \Sigma_{q^*,v_{-1}} & \Sigma_{q^*,p_K} & \Sigma_{q^*,p_V} & \Sigma_{q^*,\omega} & \Sigma_{q^*,\delta} \\ \Sigma_{k',q^*_{-1}} & \Sigma_{k',k} & \Sigma_{k',v_{-1}} & \Sigma_{k',p_K} & \Sigma_{k',p_V} & \Sigma_{k',\omega} & \Sigma_{k',\delta} \\ \Sigma_{v,q^*_{-1}} & \Sigma_{v,k} & \Sigma_{v,v_{-1}} & \Sigma_{v,p_K} & \Sigma_{v,p_V} & \Sigma_{v,\omega} & \Sigma_{v,\delta} \\ \Sigma_{p'_K,q^*_{-1}} & \Sigma_{p'_K,k} & \Sigma_{p'_K,v_{-1}} & \Sigma_{p'_K,p_K} & \Sigma_{p'_K,p_V} & \Sigma_{p'_K,\omega} & \Sigma_{p'_K,\delta} \\ \Sigma_{p'_V,q^*_{-1}} & \Sigma_{p'_V,k} & \Sigma_{p'_V,v_{-1}} & \Sigma_{p'_V,p_K} & \Sigma_{p'_V,p_V} & \Sigma_{p'_V,\omega} & \Sigma_{p'_V,\delta} \\ \Sigma_{\omega',q^*_{-1}} & \Sigma_{\omega',k} & \Sigma_{\omega',v_{-1}} & \Sigma_{\omega',p_K} & \Sigma_{\omega',p_V} & \Sigma_{\omega',\omega} & \Sigma_{\omega',\delta} \\ \Sigma_{\delta',q^*_{-1}} & \Sigma_{\delta',k} & \Sigma_{\delta',v_{-1}} & \Sigma_{\delta',p_K} & \Sigma_{\delta',p_V} & \Sigma_{\delta',\omega} & \Sigma_{\delta',\delta} \end{pmatrix}, \qquad (45)$$

$$\Gamma^2 = \begin{pmatrix} \Sigma_{q,q_{-2}} & \Sigma_{q,k_{-1}} & \Sigma_{q,v_{-2}} & \Sigma_{q,p_{K,-1}} & \Sigma_{q,p_{V,-1}} & \Sigma_{q,\omega_{-1}} & \Sigma_{q,\delta_{-1}} \\ \Sigma_{k',q_{-2}} & \Sigma_{k',k_{-1}} & \Sigma_{k',v_{-2}} & \Sigma_{k',p_{K,-1}} & \Sigma_{k',p_{V,-1}} & \Sigma_{k',\omega_{-1}} & \Sigma_{k',\delta_{-1}} \\ \Sigma_{v,q_{-2}} & \Sigma_{v,k_{-1}} & \Sigma_{v,v_{-2}} & \Sigma_{v,p_{K,-1}} & \Sigma_{v,p_{V,-1}} & \Sigma_{v,\omega_{-1}} & \Sigma_{v,\delta_{-1}} \\ \Sigma_{p'_K,q_{-2}} & \Sigma_{p'_K,k_{-1}} & \Sigma_{p'_K,v_{-2}} & \Sigma_{p'_K,p_{K,-1}} & \Sigma_{p'_K,p_{V,-1}} & \Sigma_{p'_K,\omega_{-1}} & \Sigma_{p'_K,\delta_{-1}} \\ \Sigma_{p'_V,q_{-2}} & \Sigma_{p'_V,k_{-1}} & \Sigma_{p'_V,v_{-2}} & \Sigma_{p'_V,p_{K,-1}} & \Sigma_{p'_V,p_{V,-1}} & \Sigma_{p'_V,\omega_{-1}} & \Sigma_{p'_V,\delta_{-1}} \\ \Sigma_{\omega',q_{-2}} & \Sigma_{\omega',k_{-1}} & \Sigma_{\omega',v_{-2}} & \Sigma_{\omega',p_{K,-1}} & \Sigma_{\omega',p_{V,-1}} & \Sigma_{\omega',\omega_{-1}} & \Sigma_{\omega',\delta_{-1}} \\ \Sigma_{\delta',q_{-2}} & \Sigma_{\delta',k_{-1}} & \Sigma_{\delta',v_{-2}} & \Sigma_{\delta',p_{K,-1}} & \Sigma_{\delta',p_{V,-1}} & \Sigma_{\delta',\omega_{-1}} & \Sigma_{\delta',\delta_{-1}} \end{pmatrix}, \qquad (46)$$

where we use Σ_{q^*,q^*} as shorthand for $Var(q^*)$, $\Sigma_{q^*,k'}$ for $Cov(q^*,k')$, etc.

Derivation: prediction error and proof of Proposition 1. To obtain the prediction error $\zeta = q^* - E(q^*|z)$, note that

$$\begin{pmatrix} q^* \\ k' \\ v \\ p'_V \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_{q^*} \\ \mu_{k'} \\ \mu_v \\ \mu_{p'_V} \end{pmatrix}, \begin{pmatrix} \Sigma_{q^*,q^*} & \Sigma_{q^*,k} & \Sigma_{q^*,v} & \Sigma_{q^*,p_V} \\ \dots & \Sigma_{k,k} & \Sigma_{k,v} & \Sigma_{k,p_V} \\ \dots & \dots & \Sigma_{v,v} & \Sigma_{v,p_V} \\ \dots & \dots & \dots & \Sigma_{p_V,p_V} \end{pmatrix} \right),$$

where the elements of the expectation vector and the covariance matrix are given by equations (43), (44), and (45). Normality implies that

$$E(q^*|z) = \mu_{q^*} + \begin{pmatrix} \Sigma_{q^*,k} & \Sigma_{q^*,v} & \Sigma_{q^*,p_V} \end{pmatrix} \begin{pmatrix} \Sigma_{k,k} & \Sigma_{k,v} & \Sigma_{k,p_V} \\ \dots & \Sigma_{v,v} & \Sigma_{v,p_V} \\ \dots & \dots & \Sigma_{p_V,p_V} \end{pmatrix}^{-1} \begin{pmatrix} k - \mu_{k'} \\ v - \mu_v \\ p_V - \mu_{p'_V} \end{pmatrix}.$$

We thus obtain the prediction error in closed form as

$$\zeta = q^* - \mu_{q^*} - \pi_k(k - \mu_{k'}) - \pi_v(v - \mu_v) - \pi_{p_V}(p_V - \mu_{p'_V}). \tag{47}$$

Proposition 1 is proven by evaluating equation (47).

Derivation: projection. To obtain the projection $\widehat{v} = E(v|k, p_V, z_{-1})$, note that

$$\begin{pmatrix} v \\ k \\ k_{-1} \\ v_{-1} \\ p_V \\ p_{V,-1} \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_v \\ \mu_{k'} \\ \mu_{k'} \\ \mu_v \\ \mu_{p'_V} \\ \mu_{p'_V} \end{pmatrix}, \begin{pmatrix} \Sigma_{v,v} & \Sigma_{k,v} & \Sigma_{k_{-1},v} & \Sigma_{v,v_{-1}} & \Sigma_{v,p_V} & \Sigma_{v,p_{V,-1}} \\ \dots & \Sigma_{k,k} & \Sigma_{k,k_{-1}} & \Sigma_{k,v_{-1}} & \Sigma_{k,p_V} & \Sigma_{k,p_{V,-1}} \\ \dots & \dots & \Sigma_{k_{-1},k_{-1}} & \Sigma_{k_{-1},v_{-1}} & \Sigma_{k_{-1},p_V} & \Sigma_{k_{-1},p_{V,-1}} \\ \dots & \dots & \dots & \Sigma_{v_{-1},v_{-1}} & \Sigma_{v_{-1},p_V} & \Sigma_{v_{-1},p_{V,-1}} \\ \dots & \dots & \dots & \dots & \Sigma_{p_V,p_V} & \Sigma_{p_V,p_{V,-1}} \\ \dots & \dots & \dots & \dots & \dots & \Sigma_{p_{V,-1},p_{V,-1}} \end{pmatrix}\right),$$

where the elements of the expectation vector and the covariance matrix are given by equations (43), (44), (45), and (46). Normality implies that

$$E(v|k, p_V, z_{-1}) = \mu_v + \begin{pmatrix} \Sigma_{k,v} & \Sigma_{k_{-1},v} & \Sigma_{v,v_{-1}} & \Sigma_{v,p_V} & \Sigma_{v,p_{V,-1}} \end{pmatrix}$$
$$\cdot \begin{pmatrix} \Sigma_{k,k} & \Sigma_{k,k_{-1}} & \Sigma_{k,v_{-1}} & \Sigma_{k,p_V} & \Sigma_{k,p_{V,-1}} \\ \dots & \Sigma_{k_{-1},k_{-1}} & \Sigma_{k_{-1},v_{-1}} & \Sigma_{k_{-1},p_V} & \Sigma_{k_{-1},p_{V,-1}} \\ \dots & \dots & \Sigma_{v_{-1},v_{-1}} & \Sigma_{v_{-1},p_V} & \Sigma_{v_{-1},p_{V,-1}} \\ \dots & \dots & \dots & \Sigma_{p_V,p_V} & \Sigma_{p_V,p_{V,-1}} \\ \dots & \dots & \dots & \dots & \Sigma_{p_{V,-1},p_{V,-1}} \end{pmatrix}^{-1} \begin{pmatrix} k - \mu_{k'} \\ k_{-1} - \mu_{k'} \\ v_{-1} - \mu_v \\ p_V - \mu_{p'_V} \\ p_{V,-1} - \mu_{p'_V} \end{pmatrix}.$$

We thus obtain the projection in closed form as

$$\widehat{v} = \mu_v + \phi_k(k - \mu_{k'}) + \phi_{k_{-1}}(k_{-1} - \mu_{k'}))$$
$$+\phi_{v_{-1}}(v_{-1} - \mu_v)) + \phi_{p_V}(p_V - \mu_{p'_V}) + \phi_{p_{V,-1}}(p_{V,-1} - \mu_{p'_V})). \tag{48}$$

Derivation: bias and proofs of Propositions 2 and 3. With the prediction error ζ and the projection $\widehat{v}$ in hand, we turn to the various terms in equation (16). Using equation (47) we obtain $E(k\zeta_{-1})$ as

$$E(k\zeta_{-1})$$
$$= E\left(k\left(q^*_{-1} - \mu_{q^*} - \pi_k(k_{-1} - \mu_{k'}) - \pi_v(v_{-1} - \mu_v) - \pi_{p_V}(p_{V,-1} - \mu_{p'_V})\right)\right)$$
$$= \Sigma_{q^*_{-1},k} - \pi_k\Sigma_{k,k_{-1}} - \pi_v\Sigma_{k,v_{-1}} - \pi_{p_V}\Sigma_{k,p_{V,-1}}. \tag{49}$$

Using equations (48) and (47) we next obtain $E(\widehat{v}\zeta_{-1})$ as

$$\begin{aligned}
&E(\widehat{v}\zeta_{-1})\\
-&E\Big(\Big(\mu_v + \phi_k(k-\mu_{k'}) + \phi_{k_{-1}}(k_{-1}-\mu_{k'}) + \phi_{v_{-1}}(v_{-1}-\mu_v) + \phi_{p_V}(p_V - \mu_{p'_V}) + \phi_{p_{V,-1}}(p_{V,-1}-\mu_{p'_V})\Big)\\
&\cdot\Big(q^*_{-1} - \mu_{q^*} - \pi_k(k_{-1}-\mu_{k'}) - \pi_v(v_{-1}-\mu_v) - \pi_{p_V}(p_{V,-1}-\mu_{p'_V})\Big)\Big)\\
&= \phi_k\Sigma_{q^*_{-1},k} + \phi_{k_{-1}}\Sigma_{q^*_{-1},k_{-1}} + \phi_{v_{-1}}\Sigma_{q^*_{-1},v_{-1}} + \phi_{p_V}\Sigma_{q^*_{-1},p_V} + \phi_{p_{V,-1}}\Sigma_{q^*_{-1},p_{V,-1}}\\
&-\pi_k\left(\phi_k\Sigma_{k,k_{-1}} + \phi_{k_{-1}}\Sigma_{k_{-1},k_{-1}} + \phi_{v_{-1}}\Sigma_{k_{-1},v_{-1}} + \phi_{p_V}\Sigma_{k_{-1},p_V} + \phi_{p_{V,-1}}\Sigma_{k_{-1},p_{V,-1}}\right)\\
&-\pi_v\left(\phi_k\Sigma_{k,v_{-1}} + \phi_{k_{-1}}\Sigma_{k_{-1},v_{-1}} + \phi_{v_{-1}}\Sigma_{v_{-1},v_{-1}} + \phi_{p_V}\Sigma_{v_{-1},p_V} + \phi_{p_{V,-1}}\Sigma_{v_{-1},p_{V,-1}}\right)\\
&-\pi_{p_V}\left(\phi_k\Sigma_{k,p_{V,-1}} + \phi_{k_{-1}}\Sigma_{k_{-1},p_{V,-1}} + \phi_{v_{-1}}\Sigma_{v_{-1},p_{V,-1}} + \phi_{p_V}\Sigma_{p_V,p_{V,-1}} + \phi_{p_{V,-1}}\Sigma_{p_{V,-1},p_{V,-1}}\right).
\end{aligned}$$

Direct calculation with a computer algebra system proves the following lemma:

Lemma 4 $E(\widehat{v}\zeta_{-1}) = \phi_k E(k\zeta_{-1})$.

Hence, $E(k\zeta_{-1}) = 0$ implies $E(\widehat{v}\zeta_{-1}) = 0$. Propositions 2 and 3 are therefore proven by evaluating equation (49).

Evaluating the bias outside the special cases in Propositions 1–3 requires the remaining terms in equation (16). We obtain

$$\begin{aligned}
E((k-\rho_\omega k_{-1})^2) &= E(k^2) - 2\rho_\omega E(kk_{-1}) + \rho_\omega^2 E(k_{-1}^2)\\
&= (1+\rho_\omega^2)\left(\Sigma_{k,k} + \mu_{k'}^2\right) - 2\rho_\omega\left(\Sigma_{k,k_{-1}} + \mu_{k'}^2\right)\\
E((k-\rho_\omega k_{-1})(v-\rho_\omega v_{-1})) &= E(kv) - \rho_\omega E(kv_{-1}) - \rho_\omega E(k_{-1}v) + \rho_\omega^2 E(k_{-1}v_{-1})\\
&= (1+\rho_\omega^2)\left(\Sigma_{k,v} + \mu_{k'}\mu_v\right) - \rho_\omega\left(\Sigma_{k,v_{-1}} + \mu_{k'}\mu_v\right) - \rho_\omega\left(\Sigma_{k_{-1},v} + \mu_{k'}\mu_v\right)\\
E((\widehat{v}-\rho_\omega v_{-1})(v-\rho_\omega v_{-1})) &= E(\widehat{v}v) - \rho_\omega E(\widehat{v}v_{-1}) - \rho_\omega E(vv_{-1}) + \rho_\omega^2 E(v_{-1}^2)\\
&= E(\widehat{v}v) - \rho_\omega E(\widehat{v}v_{-1}) - \rho_\omega\left(\Sigma_{v,v_{-1}} + \mu_v^2\right) + \rho_\omega^2\left(\Sigma_{v,v} + \mu_v^2\right)
\end{aligned}$$

where using equation (48)

$$\begin{aligned}
&E(\widehat{v}v)\\
&= E\left(\left(\mu_v + \phi_k(k-\mu_{k'}) + \phi_{k_{-1}}(k_{-1}-\mu_{k'}) + \phi_{v_{-1}}(v_{-1}-\mu_v) + \phi_{p_V}(p_V-\mu_{p'_V}) + \phi_{p_{V,-1}}(p_{V,-1}-\mu_{p'_V})\right)v\right)\\
&= \mu_v^2 + \phi_k\Sigma_{k,v} + \phi_{k_{-1}}\Sigma_{k_{-1},v} + \phi_{v_{-1}}\Sigma_{v,v_{-1}} + \phi_{p_V}\Sigma_{v,p_V} + \phi_{p_{V,-1}}\Sigma_{v,p_{V,-1}}
\end{aligned}$$

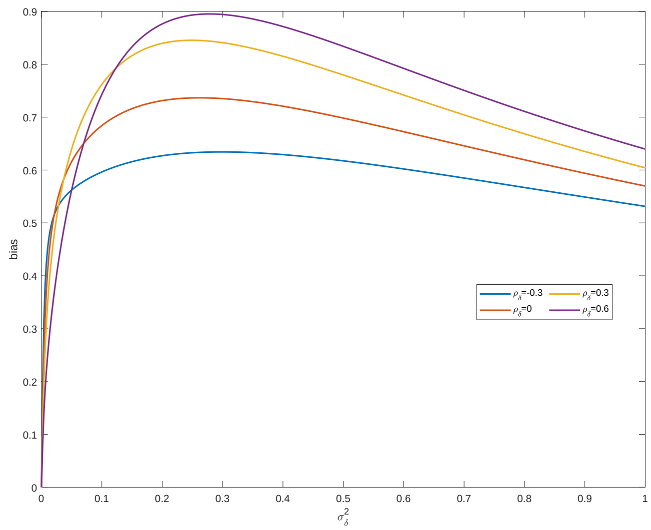

Figure 3: $bias = \frac{\widehat{\beta}_V - \beta_V}{\beta_V}$ over range of σ_δ^2 for $\rho_\delta \in \{-0.3, 0, 0.3, 0.6\}$. $\tau_0 = \tau_{p_V} = 0$, $\tau_k = \tau_\omega = \tau_\delta = 0.5$, and $\tau_{p_K} = -0.5$.

and

$$E(\widehat{v}v_{-1})$$
$$= E\left(\left(\mu_v + \phi_k(k - \mu_{k'}) + \phi_{k_{-1}}(k_{-1} - \mu_{k'}) + \phi_{v_{-1}}(v_{-1} - \mu_v) + \phi_{p_V}(p_V - \mu_{p'_V}) + \phi_{p_{V,-1}}(p_{V,-1} - \mu_{p'_V})\right) v_{-1}\right)$$
$$= \mu_v^2 + \phi_k \Sigma_{k,v_{-1}} + \phi_{k_{-1}} \Sigma_{k_{-1},v_{-1}} + \phi_{v_{-1}} \Sigma_{v_{-1},v_{-1}} + \phi_{p_V} \Sigma_{v_{-1},p_V} + \phi_{p_{V,-1}} \Sigma_{v_{-1},p_{V,-1}}.$$

Bias. Plugging the above expressions into equation (16), we obtain the bias in the estimate of the output elasticity of the variable input V in closed form. Figure 3 changes the parameterization underlying Figure 1 to $\tau_0 = \tau_{p_V} = 0$, $\tau_k = \tau_\omega = \tau_\delta = 0.5$, and $\tau_{p_K} = -0.5$. As can be seen, the bias may be enormous.

Appendix D

The ESEE is a firm-level survey of the Spanish manufacturing sector sponsored by the Ministry of Industry. At the beginning of the survey, about 5% of firms with up to 200 workers were sampled randomly by industry and size strata. All firms with more than 200 workers were included in the survey and 70% of these larger firms responded. Firms disappear over time from the sample due to either exit (shutdown or abandonment of activity) or attrition. To preserve representativeness, samples of newly created firms were added to the initial sample almost every year and some additions counterbalanced attrition.

We observe firms for a maximum of 23 years between 1990 and 2012. We restrict the sample to firms with at least three years of observations, giving a total of 3026 firms and 26977 observations. The number of firms with 3, 4,..., 23 years of data is 398, 298, 279, 278, 290, 324, 122, 111, 137, 96, 110, 66, 66, 98, 66, 40, 37, 44, 37, 42, and 87, respectively.[32]

In what follows we list the variables that we use, beginning with the variables that we take directly from the data source.

- *Revenue* (R). Value of produced goods and services computed as sales plus the variation of inventories.
- *Exports* (X). Value of exports.
- *Investment* (I). Value of current investments in equipment goods (excluding buildings, land, and financial assets) deflated by a price index of investment. The price of investment is the equipment goods component of the index of industry prices computed and published by the Spanish Ministry of Industry.
- *Capital* (K). Capital at current replacement values is computed recursively from an initial estimate and the data on investments I at $t-1$ using industry-specific depreciation rates. Capital in real terms is obtained by deflating capital at current replacement values by the price index of investment.
- *Labor* (L). 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.

[32] Table D1 in Doraszelski & Jaumandreu (2019) shows the industry labels along with their definitions in terms of the ESEE, ISIC and NACE classifications and the number of firms and observations per industry.

- *Intermediate consumption* (MB). Value of intermediate consumption (including raw materials, components, energy, and services).

- *Proportion of white collar workers* (pwc). Fraction of non-production workers.

- *Advertising* (adv). Firm expenditure in advertising.

- *R&D Expenditures* ($R\&D$). Cost of intramural R&D activities, payments for outside R&D contracts with laboratories and research centers, and payments for imported technology in the form of patent licensing or technical assistance, with the various expenditures defined according to the OECD Frascati and Oslo manuals.

- *Price of output* (P). Firm-level price index for output. Firms are asked about the price changes they made during the year in up to five separate markets in which they operate. The price index is computed as a Paasche-type index of the responses.

- *Price of labor* (P_L). Hourly wage cost computed as wage bill divided by total hours worked.

- *Price of materials* (P_M). 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 of the responses.

- *Market dynamism* (mdy). Firms are asked to assess the evolution of the main market in which they operate. The demand shifter codes the responses as 0, 0.5, and 1 for slump, stability, and expansion, respectively.

We construct a number of additional variables. We consistently subtract advertising from intermediate consumption because it is not a production input. We define variable cost as the wage bill plus the cost of intermediate consumption (minus advertising), minus the R&D expenditures and an estimate of the part of the wage bill corresponding to white collar workers. The estimation assumes that white-collar employees work the same number of hours but have an average wage 1.25 times higher. This is important to better approximate variable cost.

- *Output* (Q). Revenue deflated by the firm-specific price index of output.
- *Materials* (M). Value of intermediate consumption minus advertising deflated by the firm-specific price index of materials.
- *Variable cost* (VC). Wage bill (including social security payments) plus the cost of intermediate consumption minus advertising, R&D expenditures, and white collar pay.
- *Export status* (xst). Export status is one when the value of exports is positive and zero otherwise.

Appendix E

Let γ be the parameters estimated in equation (4) in the first step of ACF and θ the parameters estimated in equation (5) in the second step of ACF. Given the estimate $\widehat{\gamma}$, $\xi_{jt} + \varepsilon_{jt} = r_{jt}(\theta, \phi(z_{jt-1}; \widehat{\gamma}))$ in equation (5) depends on $\widehat{\phi}(z_{jt-1}) = \phi(z_{jt-1}; \widehat{\gamma})$. Stacking yields the $T_j \times 1$ vector $r_j(\theta, \phi(z_{j,-1}; \widehat{\gamma}))$, where T_j is the number of observations for firm j.

Following Wooldridge (2010), let

$$D_0 = E[w_j' r_j(\theta_0, \phi(z_{j,-1}; \widehat{\gamma})) r_j(\theta_0, \phi(z_{j,-1}; \widehat{\gamma}))' w_j]$$

be the variance of the orthogonality conditions based on the $T_j \times Q$ matrix of instruments w_j in the second step of ACF, evaluated at the true value of θ. Expanding $r_j(\cdot)$ around the true value of γ yields $r_j(\theta_0, \phi(z_{j,-1}; \widehat{\gamma})) \approx r_j(\theta_0, \phi(z_{j,-1}; \gamma_0)) + \frac{\partial r_j}{\partial \phi} \nabla_\gamma \phi(z_{j,-1}; \gamma_0)(\widehat{\gamma} - \gamma_0)$. Since γ is estimated by OLS, we use $(\widehat{\gamma} - \gamma_0) = \sum_j (f(z_j)' f(z_j))^{-1} f(z_j)' \varepsilon_j$, where $f(z_j)$ are the regressors in the first step of ACF, and replace $r_j(\cdot)$ by

$$\widetilde{r}_j(\theta_0, \gamma_0, \varepsilon_j) = r_j(\theta_0, \phi(z_{j,-1}; \gamma_0)) + \frac{\partial r_j}{\partial \phi} \nabla_\gamma \phi(z_{j,-1}; \gamma_0) \textstyle\sum_j (f(z_j)' f(z_j))^{-1} f(z_j)' \varepsilon_j.$$

Replacing the true values of θ, γ, and ε_j by their estimates, we estimate D_0 as $\widehat{D} = \frac{1}{N} \sum_j w_j' \widetilde{r}_j(\widehat{\theta}, \widehat{\gamma}, \widehat{\varepsilon}_j) \widetilde{r}_j(\widehat{\theta}, \widehat{\gamma}, \widehat{\varepsilon}_j)' w_j$. Next, we use $\widehat{D}$ in the usual sandwich formula for the asymptotic variance of the estimated parameters θ and in the optimal weighting matrix to compute the Sargan test.

References

Abito, M. (2022), Estimating production functions with unobserved time-invariant heterogeneity and time-varying productivity shocks, Working paper, Ohio State University, Columbus.

Ackerberg, D. (2020), Timing assumptions and efficiency: Empirical evidence in a production function context, Working paper, University of Texas, Austin.

Ackerberg, D., Benkard, L., Berry, S. & Pakes, A. (2007), Econometric tools for analyzing market outcomes, *in* J. Heckman & E. Leamer, eds, 'Handbook of Econometrics', Vol. 6A, North-Holland, Amsterdam, pp. 4171–4276.

Ackerberg, D., Caves, K. & Frazer, G. (2015), 'Identification properties of recent production function estimators', *Econometrica* **83**(6), 2411–2451.

Adda, J. & Cooper, R. (2003), *Dynamic economics: Quantitative methods and applications*, MIT Press, Cambridge.

Arellano, M. & Bond, S. (1991), 'Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations', *Review of Economic Studies* **58**, 277–297.

Arellano, M. & Bover, O. (1995), 'Another look at instrumental variable estimation of error component models', *Journal of Econometrics* **68**, 29–52.

Autor, D., Dorn, D., Katz, L., Patterson, C. & Van Reenen, J. (2020), 'The fall of the labor share and the rise of superstar firms', *Quarterly Journal of Economics* **135**(2), 645–709.

Azzam, A., Jaumandreu, J. & Lopez, R. (2023), Input and output market power with non-neutral productivity: Livestock and labor in meatpacking, Working paper no. 17785, CEPR, London.

Bain, J. (1951), 'Relation of profit rate to industry concentration: American manufacturing, 1936-1940', *Quarterly Journal of Economics* **65**, 293–324.

Baker, J. & Bresnahan, T. (1985), 'The gains from merger or collusion in product-differentiated industries', *Journal of Industrial Economics* **33**, 59–76.

Bernstein, J. & Mohnen, P. (1991), 'Price-cost margins, exports and productivity growth: with an application to Canadian industries', *Canadian Journal of Economics* **24**(3), 638–659.

Berry, S., Gaynor, M. & Scott-Morton, F. (2020), 'Do increasing markups matter? Lessons from empirical industrial organization', *Journal of Economic Perspectives* **33**(3), 44–68.

Berry, S., Levinsohn, J. & Pakes, A. (1995), 'Automobile prices in market equilibrium', *Econometrica* **63**(4), 841–890.

Biondi, F. (2022), Firm productivity and derived factor demand under variable markups, Working paper, Katholieke Universiteit Leuven, Leuven.

Blum, B., Claro, S., Horstmann, I. & Rivers, D. (2023), 'The ABCs of firm heterogeneity when firms sort into markets: The case of exporters', *Journal of Political Economy* **forthcoming**.

Blundell, R. & Bond, S. (1998), 'Initial conditions and moment restrictions in dynamic panel data models', *Journal of Econometrics* **87**, 115–143.

Blundell, R. & Bond, S. (2000), 'GMM estimation with persistent panel data: An application to production functions', *Econometric Reviews* **19**, 321–340.

Bond, S., Hashemi, G., Kaplan, G. & Zoch, P. (2021), 'Some unpleasant markup arithmetic: Production function elasticities and their estimation from production data', *Journal of Monetary Economics* **121**(C), 1–14.

Brand, J. (2020), Estimating productivity and markups under imperfect competition, Working paper, University of Texas, Austin.

Brandt, L., Van Biesebroeck, J., Wang, L. & Zhang, Y. (2017), 'WTO accession and performance of Chinese manufacturing firms', *American Economic Review* **107**(9), 2784–2820.

Brandt, L., Van Biesebroeck, J., Wang, L. & Zhang, Y. (2019), 'WTO accession and performance of Chinese manufacturing firms: Corrigendum', *American Economic Review* **109**(4), 1616–1621.

Bughin, J. (1996), 'Capacity constraints and export performance: Theory and evidence from Belgian manufacturing', *Journal of Industrial Economics* **44**(3), 187–204.

Collard-Wexler, A. & De Loecker, J. (2020), Production function estimation and capital measurement error, Working paper, Duke University, Durham.

Das, S., Roberts, M. & Tybout, J. (2007), 'Market entry costs, producer heterogeneity, and export dynamics', *Econometrica* **75**(3), 837–873.

De Loecker, J. & Eeckhout, J. (2018), Global market power, Working paper no. 24768, NBER, Cambridge.

De Loecker, J., Eeckhout, J. & Unger, G. (2018), The rise of market power and the macroeconomic implications, Working paper, Katholieke Universiteit Leuven, Leuven.

De Loecker, J., Eeckhout, J. & Unger, G. (2020), 'The rise of market power and the macroeconomic implications', *Quarterly Journal of Economics* **135**(2), 561–644.

De Loecker, J., Goldberg, P., Khandelwal, A. & Pavcnik, N. (2016), 'Prices, markups and trade reform', *Econometrica* **84**(2), 445–510.

De Loecker, J. & Scott, P. (2016), Estimating market power: Evidence from the US brewing industry, Working paper no. 22957, NBER, Cambridge.

De Loecker, J. & Syverson, C. (2021), An industrial organization perspective on productivity, *in* K. Ho, A. Hortacsu & A. Lizzeri, eds, 'Handbook of Industrial Organization', Vol. 4, North-Holland, Amsterdam, pp. 141–223.

De Loecker, J. & Warzynski, F. (2012), 'Markups and firm-level export status', *American Economic Review* **102**(6), 2437–2471.

Demirer, M. (2019), Production function estimation with imperfect proxies, Working paper, MIT, Cambridge.

Demirer, M. (2020), Production function estimation with factor-augmenting technology: An application to markups, Working paper, MIT, Cambridge.

Doraszelski, U. & Jaumandreu, J. (2018), 'Measuring the bias of technological change', *Journal of Political Economy* **126**(3), 1027–1084.

Doraszelski, U. & Jaumandreu, J. (2019), Using cost minimization to estimate markups, Working paper no. 14114, CEPR, London.

Foster, L., Haltiwanger, J. & Syverson, C. (2008), 'Reallocation, firm turnover, and efficiency: Selection on productivity or profitability?', *American Economic Review* **98**(1), 394–425.

Hall, R. (1986), 'Market structure and macroeconomic fluctuations', *Brookings Papers on Economic Activity: Macroecomomics* **17**(2), 285–322.

Hausman, J., Newey, W., Ichimura, H. & Powell, J. (1991), 'Identification and estimation of polynomial errors-in-variables models', *Journal of Econometrics* **50**, 273–295.

Hsieh, C. & Klenow, P. (2009), 'Misallocation and manufacturing TFP in China and India', *Quarterly Journal of Economics* **124**(4), 1403–1448.

Hu, Y., Huang, G. & Sasaki, Y. (2020), 'Estimating production functions with robustness against errors in the proxy variables', *Journal of Econometrics* **215**(2), 375–398.

Hu, Y. & Schennach, S. (2008), 'Instrumental variable treatment of nonclassical measurement error models', *Econometrica* **76**(1), 195–216.

Jaumandreu, J. & Yin, H. (2018), Cost and product advantages: Evidence from Chinese manufacturing firms, Working paper no. 11862, CEPR, London.

Kim, K., Petrin, A. & Song, S. (2016), 'Estimating production functions with control functions when capital is measured with error', *Journal of Econometrics* **190**(2), 267–279.

Klette, T. (1999), 'Market power, scale economies and productivity: Estimates from a panel of establishment data', *Journal of Industrial Economics* **67**(4), 451–476.

Levinsohn, J. & Petrin, A. (2003), 'Estimating production functions using inputs to control for unobservables', *Review of Economic Studies* **70**(2), 317–341.

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer, Berlin.

Moreno, L. & Rodriguez, D. (2004), 'Domestic and foreign price/marginal-cost margins: An application to Spanish manufacturing firms', *Review of International Economics* **12**(1), 60–80.

Mundlak, Y. & Hoch, I. (1965), 'Consequences of alternative specifications in estimation of Cobb-Douglas production functions', *Econometrica* **33**(4), 814–828.

Olley, S. & Pakes, A. (1996), 'The dynamics of productivity in the telecommunications industry', *Econometrica* **64**(6), 1263–1297.

Pakes, A. (1994), Dynamic structural models, problems, and prospects: mixed continuous discrete controls and market interactions, *in* C. Sims, ed., 'Advances in econometrics: Sixth World Congress', Vol. 2, Cambridge University Press, Cambridge.

Ponder, M. (2021), Production functions without control functions: Estimating input elasticities when monotonicity does not hold, Working paper, University of Minnesota, Minneapolis.

Raval, D. (2019), 'The micro elasticity of substitution and non-neutral technology', *Rand Journal of Economics* **50**(1), 147–167.

Raval, D. (2022), 'Testing the production function approach to markup estimation', *Review of Economic Studies* **forthcoming**.

Rubens, M. (2023), 'Market structure, oligopsony power, and productivity', *American Economic Review* **113**(9), 2382–2410.

Syverson, C. (2001), Market structure and productivity, Ph.d. dissertation, University of Maryland, College Park.

Syverson, C. (2004), 'Market structure and productivity: A concrete example', *Journal of Political Economy* **112**(6), 1181–1222.

Van Biesebroeck, J. (2007), 'Robustness of productivity estimates', *Journal of Industrial Economics* **55**(3), 529–569.

Wooldridge, J. (2010), *Econometric analysis of cross section and panel data*, 2nd edn, MIT Press, Cambridge.

Yeh, C., Macaluso, C. & Hershbein, B. (2022), 'Monopsony in the US labor market', *American Economic Review* **112**(7), 2099–2138.

Zhang, H. (2019), 'Non-neutral technology, firm heterogeneity, and labor demand', *Journal of Development Economics* **140**, 145–168.