The Hitchhiker's Guide to Markup Estimation*

Maarten De Ridder† Basile Grassi‡ Giovanni Morzenti§

Preliminary Draft
July 8 2021

Abstract

How do estimates of firm-level markups that rely on production function estimations depend on common data limitations? With a tractable analytical framework, simulation from a quantitative model, and firm-level administrative production and pricing data, we study biases due to the use of revenue instead of quantity, and due to production function misspecification. Estimates from revenue mismeasure the level of markups, but do contain useful information about true markups. Conversely, misspecified production functions have little effect on the estimated average markup but reduce their information content. Finally, revenue and quantity markups display similar correlations with variables such as profitability and market share in our data.

Keywords: Macroeconomics, Production Functions, Markups, Competition

*We thank Marco Panunzi for his great research assistance. We would like to thank Jêrome Adda, Ariel Burstein, Vasco M. Carvalho, Francesco Decarolis, Marco Maffezzoli, Massimiliano Marcellino, Gianmarco Ottaviano, and, Ariell Reshef for useful comments. We also thank participants at the Society for Economic Dynamics, Bocconi, LMU and the Paris School of Economics for useful comments and suggestions. We thank Isabelle Mejean for help with the French data. This work is supported by a public grant overseen by the French National Research Agency (ANR) as part of the 'Investissements d'Avenir' program (reference: ANR-10-EQPX-1, CASD).

†London School of Economics. m.c.de-ridder@lse.ac.uk

‡Bocconi University, CEPR and IGIER. basile.grassi@unibocconi.it

§Bocconi University. giovanni.morzenti@phd.unibocconi.it

1 Introduction

The markup is a key variable in macroeconomics. Markups above one mean that firms charge prices above marginal costs, which reduces demand for goods. This negatively affects demand for production factors and creates a gap between factor prices and marginal productivities. If markups differ across firms there is a further effect: high-markup firms produce inefficiently little compared to low-markup firms, causing output in the economy to fall below the efficient frontier and depressing aggregate total factor productivity (e.g. Hsieh and Klenow 2009). Conversely, higher markups can represent the return to intangible investments and fixed costs, and can incentivize firms to innovate and drive aggregate productivity growth (e.g. Aghion et al. 2005). Whether this is the case has become a particularly relevant question recently, because new evidence suggests that both the level and the dispersion of markups have increased steadily over the last 40 years (e.g. De Loecker et al. 2020).

To analyze the macroeconomic effects of markups, it is first needed to measure them. This paper uses a combination of theory, simulations and detailed firm-level administrative data on production and prices to assess the performance of empirical markup estimation methods. Measuring markups is not trivial as the true markup is the ratio of price over marginal costs. The former is only available in a few datasets, the latter in virtually none. Recent work therefore uses the ratio estimator of the markup, which was first derived in Hall (1986, 1988). He derives that markups can be found by multiplying the output elasticity of some variable input, that firms set without adjustment costs, with the ratio of revenue over the firm's spending on that variable input. Markups are therefore straightforward to estimate with data on revenue and input spending from the income statement, if the researcher knows the parameters of the production function.

The purpose of this paper is twofold. First, we provide an accessible introduction to the two-stage procedure that is commonly used to obtain production function parameters to estimate markups. Estimating a production function is difficult because firm-level productivity is unobserved. Productivity directly affects output and indirectly affects inputs (because productive firms charge lower prices and therefore sell more), such that a least squares estimation of a parametric production function is biased (e.g. Klette and Griliches 1996). To solve this, many papers after De Loecker and Warzynski (2012) have estimated markups using production function elasticities from the Ackerberg et al. (2015) two-stage iterative GMM procedure. The iterative GMM procedure involves a first-stage regression to purge observed firm output of measurement error. The second stage involves a regression to identify the production function, imposing structure on the productivity process to identify the true parameters. We provide an intuitive discussion in a simplified analytical framework of the conditions under which a researcher is able to consistently estimate the elasticities of the production function, and therefore the markup, with this procedure.

In the body of the paper, we scrutinize three prominent critiques of this procedure in settings where firms have market power. The first critique is that researchers often use revenue data as a measure of output to estimate production functions. Because price-setting firms face downward-sloping demand curves, firms that increase production need to lower prices to attain sufficient demand. Consequently, estimates of the elasticities for inputs from a revenue production function are lower than those from a quantity production function. Bond et al. (2020) show that this causes a bias in the markup estimates severe enough to make the estimates uninformative of the true markup. Yet, the scarcity of (accessible) data on firm-level prices means that the majority of researchers are limited to the analysis of data on revenues. The Bond et al. (2020)-critique therefore has the potential to seriously limit future analysis of firm-level markups. The second critique we discuss is on the assumption of a Cobb-Douglas functional form for the production function, when the true production function is more complex. The Cobb-Douglas production function is commonly assumed when estimated markups because it allows the researcher to control for errors in the production function estimation through sector fixed effects.[1] The third critique we discuss is on two-stage GMM procedure in settings where firms have market power. The procedure involves a first-stage regression to purge a firm's output from measurement error. Recent work, including Bond et al. (2020) and Doraszelski and Jaumandreu (2020), shows that this purging requires the researcher to observe marginal costs if firms have market power, which is usually unfeasible.

While all of these critiques have merit, we show that there are strong practical differences in the extent to which estimates of firm-level markups are affected. We do so in two ways. We first estimate the production function on a sample of simulated firms. The firms compete oligopolistically as in Atkeson and Burstein (2008) and therefore have endogenously heterogeneous markups. This allows us to compare markup estimates and the true parameters of the production function with estimates from the Ackerberg et al. (2015) procedure and the Hall (1986) ratio-estimator. We then compare markup estimates in different specifications using firm-level data on the universe of French manufacturing firms with at least 20 employees. The dataset contains balance sheet and income statement data for 2009 to 2018, as well as data on the unit values of the products they sell. This enables us to empirically assess the degree to which markup estimates from revenue data differ from markup estimates from quantity data.

When comparing the estimates of markups derived from revenue and quantity production functions, we find that revenue-based markups are informative of quantity-based markups. Bond et al. (2020) show that if markups are calculated from an input's revenue elasticity, the resulting markup should always equal 1. This is because price-setting firms set markups as a

[1] If firms have the same production functions within industries, an analysis of Cobb-Douglas log markups with industry fixed effects is unaffected by bias in production function estimates (e.g. Peters 2020, Crouzet and Eberly 2019, Meier and Reinelt 2020). Cobb Douglas is therefore frequently assumed in response to the Bond et al. (2020) and Doraszelski and Jaumandreu (2020) critiques.

function of the price elasticity of demand. We confirm that this critique holds *on average*, in the sense that the average level of revenue-based markups is not informative of the true average markup. Average revenue markups equal one if firms do not face aggregate demand shocks, though the average can fall short, exceed or be equal to the true markup in richer settings. We further show that markup *dispersion*, both in the cross-section and over time, is estimated reasonably well. This is because revenue-production functions do not measure the exact elasticity of revenues with respect to an input when a functional form, such as Cobb Douglas or translog, is assumed. We find a positive pairwise correlation of 0.34 for our preferred revenue and quantity-based log markup estimates, rising to 0.61 in log differences. We furthermore find that the firm-level correlation of markup estimates with profitability, the labor share and market share is similar for quantity-based markups and revenue-based markups. We conclude that revenue-based markup estimates do contain meaningful information about true markups.

We analyze the effect of assuming a Cobb-Douglas production function by comparing the resulting markups with estimates from a translog production function. The latter is a first-order log approximation of many more complex production functions, including the CES, and nests Cobb-Douglas as a special case. We document significant drawbacks to assuming that production is Cobb Douglas production when true production does not. While our simulations show that average markups are well-estimated, dispersion is not. For the French data, we find that estimated dispersion in markups under the Cobb-Douglas assumption is more than twice the dispersion found when assuming the less restrictive translog. In a back-of-the-envelope exercise, we show that researchers that attempt to measure the welfare costs of markups using a Cobb-Douglas production function would overestimate these costs by 126% in the French data.

We find that estimates of the markup are mostly robust to misspecification of the first stage in the two-stage iterative GMM procedure. The first stage is valid if it purges measurement error from the observed firm output. In practice, this is done by running a regression of output on a third-degree polynomial of observed inputs and a set of control variables. The residuals identify the measurement error as long as unobserved total factor productivity is correctly accounted for by the polynomial and control variables. If firms are price-setters, one of those controls should be the log marginal cost. As marginal costs are usually unobserved, a researcher could instead include log prices and log markups. While markups are unobserved, a researcher could instead add a variable that closely correlates with markups, such as market share (e.g. De Loecker et al. 2020). We compare markup estimates from three specifications: one where we do not purge measurement error, one where we only include a third-degree polynomial of inputs in the first stage, and one where we additionally include price and market share. We show in simulations, where we know the 'true' measurement error, the most complete full stage correctly identifies the errors. We further show that the first stage has minor effects on the estimates of the production function and markups, both in simulations and in data. We conclude the Ackerberg et al. (2015) procedure can feasibly be used to estimate firm-level markups.

Related literature Our paper builds on a significant literature that estimates production functions at the firm level. A simple regression of a set of (log) inputs on a firm's (log) output does not identify the production function elasticities because of unobserved differences in productivity across firms. Productivity directly affects output and indirectly affects inputs (e.g. because productive firms charge lower prices and therefore sell more), such that a least squares estimation of a parametric production function is biased (e.g. Klette and Griliches 1996). Our analysis is in the spirit of the seminal work by Olley and Pakes (1996) and Levinsohn and Petrin (2003) by using a proxy regression to control unobservable productivity. In line with Ackerberg et al. (2015) we use materials as a variable input to proxy for unobservable productivity, under the assumption that demand for materials is monotonic in the latter.

We focus on the feasibility of using estimates of production function elasticity to estimate markups at the firm level. This technique was pioneered by De Loecker and Warzynski (2012) to show that exporting firms have higher markups than non-exporters. De Loecker et al. (2016) extend their methodology to multi-product firms.[2] De Loecker et al. (2020) applies the methodology to listed U.S. firms to show that estimated markups have increased sharply between 1980 and 2015, a result confirmed for other countries by Díez et al. (2019) and that has sparked a rich discussion on the feasibility of the De Loecker and Warzynski (2012) methodology on accounting data (e.g., Traina 2018, Basu 2019, Syverson 2019). Baqaee and Farhi (2019) note that the rise of markups is driven by a reallocation of activity towards high-markup firms, in line with evidence of reallocation towards low-labor share firms in Autor et al. (2020) and Kehrig and Vincent (2021). Hershbein et al. (2021) and Morlacco (2019) note that markup estimates are biased when firms have market power on the market for the flexible input.[3]

Estimating the production function when firms have market power is particularly challenging. As pointed out by Doraszelski and Jaumandreu (2019) and Brand (2019) one of the identifying assumptions in the procedure by Ackerberg et al. (2015) is that the demand function by firms is not affected by unobservables other than productivity.[4] This is true for cases such as perfect or monopolistic competition but not necessarily for the case of oligopolistic competition. We analyze the bias arising from improperly accounting for demand (and therefore markup) heterogeneity in the estimation of the production function elasticities and markups.[5]

[2]Because most datasets do not provide input allocations across products that firms produce, De Loecker et al. (2016) estimate production functions at the product level using data on single-product firms. They then estimate markups for multi-product firms with the estimated elasticities. We also rely on product-level data for prices and quantities, but aggregate this to the firm level when estimating markups.

[3]A number of recent papers deploys markup estimates using the De Loecker and Warzynski (2012) methodology in specific applications. Burstein et al. (2020) show that markups in French data are either cyclical or countercyclical depending on the level of aggregation that is considered. Meier and Reinelt (2020) add that markups become more dispersed after monetary policy shocks, negatively affecting total factor productivity. Calligaris et al. (2018) find that markups have increased most in sectors with high digitization, a result confirmed at the firm level by De Ridder (2019). Crouzet and Eberly (2019) find a relationship between markups and a firm's intangible investment share.

[4]The bias arising from a violation of this assumption, in particular on correlations between markups and demand determinants, is analyzed in Doraszelski and Jaumandreu (2020).

[5]Our estimation also requires sufficient variation in input prices for the variable input to allow separate identifi-

Our paper adds to the broader literature that on the consequences of market power. Karabarbounis and Neiman (2014) find that the share of labor has fallen in most advanced economies over the last decades.[6] Barkai (2017) adds that, when accounting for the falling costs of capital, the capital share of income has also declined, leaving a rise in the profit share as the residual.[7] Increasing markups have been linked to low investments and a lack of entry (e.g. Gutiérrez and Philippon 2017, Eggertsson et al. 2018, Gutiérrez and Philippon 2018), low productivity growth (e.g. Aghion et al. 2019, De Ridder 2019),[8] and industry concentration (e.g. Grullon et al. 2019, Autor et al. 2020). The effects of markup dispersion on welfare through misallocation have been quantified in (e.g.) Baqaee and Farhi (2019), Edmond et al. (2018), and Peters (2020).[9]

Outline The remainder of this paper proceeds as follows. Section 2 outlines an analytical framework to understand the conditions under which a firm-level production function are consistently estimated, potential biases are discussed in Section 3. In Section 4 we introduce our dataset with production and pricing data for French manufacturing firms. Section 5 presents simulations that compare the performance of our various markup estimates in a setting where true markups and production functions are known. In section 6 we compare the empirical properties of markup estimates from various specifications. Section 7 concludes.

2 Analytical Framework

In this section, we discuss the properties of the two-stage Generalized Method of Moments (GMM) estimator of the production function. The first stage purges a firm's observed output from measurement error. The second stage then consists of a GMM instrumental variable estimation to estimate the actual production function on purged output. For this section, we assume that the researcher observes prices and quantities. We start by describing the second stage of the GMM estimator in the simplest possible environment where output is log-linear in a single production factor and productivity is idiosyncratic. We then extend this simple model by introducing persistent productivity, multiple production factors and more sophisticated production functions. Finally, we discuss measurement error and its purging in the first stage of the procedure, in particular when firms have market power.

cation of the variable input and productivity (e.g. Blundell and Bond 2000, Gandhi et al. 2020).

[6] Gutiérrez and Piton (2020) note that, outside of North America, the decline in the labor share is not present outside of the housing sector.

[7] Neiman and Vavra (2019) note that unmeasured inputs would also appear as a rise in profits from this calculation, and therefore label the residual of national income after labor and capital payments 'factorless income'.

[8] Cavenaile et al. (2019) note that the rise of markups also incentives firms to invest in R&D.

[9] Bornstein (2018) show that consumer demand has become less sensitive to price changes, a trend that might be driven by aging, while Neiman and Vavra (2019) consumption baskets are increasingly narrow.

2.1 The Baseline Estimator

To understand the basic intuition behind the identification of production functions, consider the case where firms produce their output Y_{it} using one input V_{it}. Both output and the input are observed without measurement error. Firms are subject to total factor productivity (TFP) shocks, denoted ω_{it} in logs, that are unobserved to the econometrician. All firms minimize costs and share the following Cobb-Douglas production function:

$$y_{it} = \alpha v_{it} + \omega_{it},$$

where the parameter α is the true output elasticity of v_{it} and where lower-caps denote log-deviations from the mean.[10] We assume that the input, v_{it}, is variable and static, that is, the choice of this inputs usage is entirely determined within the period and is not subject to any adjustment cost or information frictions. We further assume that the productivity ω_{it} is independently and identically distributed (i.i.d.) across time and firms. Because v_{it} is a flexible input, we can use it to calculate markups along the ratio estimator. Denoting the firm's price by P_{it} and the factor price of V_{it} by W_t, the true markup is given by $\mu_{it} = \alpha(P_{it}Y_{it})/(W_t V_{it})$.

To calculate these markups we must estimate the output elasticity α. A least-square regression of input v_{it} on output y_{it} is likely to be biased, because unobserved productivity ω_{it} serves as a residual and is likely to affect a firm's choice of v_{it}, violating orthogonality. The solution is to construct a GMM estimator of α using the past value of input v_{it-1} as an instrument for v_{it}:

Definition 1 *(GMM estimator) the GMM estimator is* $\widehat{\alpha} \in \mathbb{R}$ *such that the moment* $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right]$ *is equal to zero where* $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha}) v_{it} + \omega_{it}$ *.*[11]

Inserting the production function into the moment condition yields:

$$\mathbb{E}\left[((\alpha - \widehat{\alpha}) v_{it} + \omega_{it}) v_{it-1}\right] = (\alpha - \widehat{\alpha})\mathbb{E}\left[v_{it} v_{it-1}\right] = 0,$$

which uses $\mathbb{E}\left[\omega_{it} v_{it-1}\right] = 0$. It is clear that the condition above is only satisfied when $\alpha = \widehat{\alpha}$ or when $\mathbb{E}\left[v_{it} v_{it-1}\right] = 0$. Hence as long as there is a non-zero correlation between v_{it} and v_{it-1}, the estimator $\widehat{\alpha}$ is asymptotically unbiased with a unique value equal to the true elasticity α.

What assures that there is a non-zero correlation between v_{it} and v_{it-1}, or in other words, that the lagged variable input is a relevant instrument? We have assumed that productivity is i.i.d. and thus have no persistence, which means that persistence in the variable input must come from some other source. The first order condition for a cost-minimizing firm's v_{it} reads:

[10] To be precise, $x_{it} = \log X_{it} - \mathbb{E}\left[\log X_{it}\right]$ where $\mathbb{E}\left[\log X_{it}\right]$ is the limit of the empirical average across observations. This normalization allows to get rid of any constant in the production function and ensures ω_{it} has mean zero.

[11] In the above definition, the expectation operator $\mathbb{E}$ denotes the limit of the empirical average across observations. We therefore study the asymptotic properties of the GMM estimator, which allows us to keep the argument as transparent as possible. In appendix A.1, we first derive the estimator for finite sample before deducing its asymptotic variance.

$$v_{it} = \frac{1}{1-\alpha} \left(\omega_{it} + mc_{it} - w_t \right). \tag{1}$$

It follows that the persistence of v_{it} has to come from either persistence in the input price w_t or in log marginal cost mc_{it}. Note that marginal cost is equal to P_{it}/μ_{it} and is determined in equilibrium by the demand system and the competition games among firms. This argument is not unlike Gandhi et al. (2020) which point to the need of persistence in the variable inputs independently of productivity persistence for the identification of production function. We collect these results in the following proposition.

Proposition 1 *The GMM estimator is identified and asymptotically unbiased, that is,* $\widehat{\alpha} = \alpha$.

In appendix A.1, we derive the asymptotic variance of the above defined GMM estimator which satisfies $\mathbb{V}ar\left[\widehat{\alpha}\right] \sim \frac{\mathbb{E}[\omega_{it}^2]\mathbb{E}\left[v_{it-1}^2\right]}{n\mathbb{E}[v_{it}v_{it-1}]^2}$. The precision of this estimator is then proportional to the variance of productivity, $\mathbb{E}\left[\omega_{it}^2\right]$, and increases in the power of the instrument $\mathbb{E}[v_{it}v_{it-1}]^2$.

To conclude, the GMM estimator is identified and asymptotically unbiased. In the next subsection, we show that this results holds when we relax the simplifying assumptions of this model. For a first reading, the reader can skip the next section.

2.2 Extensions

We now show that the consistency of the GMM estimator is robust to several extensions that are common in practical applications. We study (i) the case of translog production function, (ii) the case of several inputs, and (iii) the case of AR(1) productivity. We discuss the case with all these extensions together in the appendix.

2.2.1 Translog Production Function

We first ease the assumption that output is log-linear by replacing the Cobb-Douglas production function with a translog specification:

$$y_{it} = \alpha v_{it} + \beta v_{it}^2 + \omega_{it}.$$

The other assumptions are unchanged. Our aim is to identify the parameters α and β, to be able to calculate the size-dependent output elasticity of the variable input for the calculation of the true markup $\mu_{it} = (\alpha + 2\beta v_{it})(P_{it}Y_{it})/(W_t V_{it})$. The least-squares estimation of the production function suffers from the same bias as before, which we address by instrumenting v_{it} and v_{it}^2 by their respective lags. Econometrically, estimating the more sophisticated translog production is therefore simply akin to estimating a multivariate GMM regression with instrumental variables. Formally, we define the estimator as:

Definition 2 *The GMM estimator is a pair* $(\widehat{\alpha}, \widehat{\beta})$ *such that* $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right] = 0$ *and* $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}^2\right] = 0$ *where* $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} - \widehat{\beta} v_{it}^2 = (\alpha - \widehat{\alpha}) v_{it} + (\beta - \widehat{\beta}) v_{it}^2 + \omega_{it}$.

It is again straightforward to solve for the estimator $(\widehat{\alpha}, \widehat{\beta})$ in our parsimonious setting. It involves solving the system of linear equations implied by the moment conditions:

$$\begin{array}{c} \mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right] = 0 \\ \mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}^2\right] = 0 \end{array} \iff \begin{array}{c} (\alpha - \widehat{\alpha})\mathbb{E}[v_{it} v_{it-1}] + (\beta - \widehat{\beta})\mathbb{E}[v_{it}^2 v_{it-1}] = 0 \\ (\alpha - \widehat{\alpha})\mathbb{E}[v_{it} v_{it-1}^2] + (\beta - \widehat{\beta})\mathbb{E}[v_{it}^2 v_{it-1}^2] = 0 \end{array}.$$

This system can be rewritten in matrix form with $V(B - \widehat{B}) = 0$ where

$$B - \widehat{B} = \begin{pmatrix} \alpha - \widehat{\alpha} \\ \beta - \widehat{\beta} \end{pmatrix} \qquad \text{and} \qquad V = \begin{pmatrix} \mathbb{E}[v_{it} v_{it-1}] & \mathbb{E}[v_{it}^2 v_{it-1}] \\ \mathbb{E}[v_{it} v_{it-1}^2] & \mathbb{E}[v_{it}^2 v_{it-1}^2] \end{pmatrix}.$$

As long as the determinant of V is not zero, the GMM estimator on translog is identified and asymptotically unbiased such that $\widehat{\alpha} = \alpha$ and $\widehat{\beta} = \beta$. This is the case as long as v_{it} and its square are not colinear and when the lagged values of v_{it} and v_{it}^2 are relevant instruments.

2.2.2 Several Inputs

In the next extension, we assume that firms produces with two inputs, a variable input v_{it} and another input k_{it}. We assume that the additional input is, in the terminology of the production function literature, dynamic. This means that firms face adjustment costs and other intertemporal constraints when setting k_{it}, that lead firms to choose k_{it} before observing contemporaneous productivity. The production function in logs reads $y_{it} = \alpha v_{it} + \beta k_{it} + \omega_{it}$ and we are interested in estimating the parameters (α,β). Because k_{it} is set before productivity is observed, we only need to instrument the variable input with its lag. The estimation is therefore akin to a GMM regression with one endogenous and one exogenous variable. The estimator is:

Definition 3 *The GMM estimator is a pair* $(\widehat{\alpha}, \widehat{\beta})$ *such that* $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right] = 0$ *and* $\mathbb{E}\left[\widehat{\omega}_{it} k_{it-1}\right] = 0$ *where* $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} - \widehat{\beta} k_{it} = (\alpha - \widehat{\alpha}) v_{it} + (\beta - \widehat{\beta}) k_{it} + \omega_{it}$.

Solving for the estimator, $(\widehat{\alpha}, \widehat{\beta})$, implies solving for the following system of equations defined by the moment conditions:

$$\begin{array}{c} \mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right] = 0 \\ \mathbb{E}\left[\widehat{\omega}_{it} k_{it-1}\right] = 0 \end{array} \iff \begin{array}{c} (\alpha - \widehat{\alpha})\mathbb{E}[v_{it} v_{it-1}] + (\beta - \widehat{\beta})\mathbb{E}[k_{it} v_{it-1}] = 0 \\ (\alpha - \widehat{\alpha})\mathbb{E}[v_{it} k_{it-1}] + (\beta - \widehat{\beta})\mathbb{E}[k_{it} k_{it-1}] = 0 \end{array}.$$

This system can be rewritten in matrix form with $V(B - \widehat{B}) = 0$ where

$$B - \widehat{B} = \begin{pmatrix} \alpha - \widehat{\alpha} \\ \beta - \widehat{\beta} \end{pmatrix} \qquad \text{and} \qquad V = \begin{pmatrix} \mathbb{E}[v_{it} v_{it-1}] & \mathbb{E}[k_{it} v_{it-1}] \\ \mathbb{E}[v_{it} k_{it-1}] & \mathbb{E}[k_{it} k_{it-1}] \end{pmatrix}.$$

Note that if the input k_{it} is perfectly correlated with the variable inputs v_{it} (and hence also variable), the matrix V won't be of full rank leading to non-identification of the estimator $(\widehat{\alpha}, \widehat{\beta})$. However, if k_{it} is not variable, it is thus not perfectly correlated with v_{it} and the determinant of V can be different from zero. As long as the determinant of V is not zero, the GMM estimator is identified and asymptotically unbiased such that $\widehat{\alpha} = \alpha$ and $\widehat{\beta} = \beta$.

2.2.3 Persistent Productivity

In the final extension, we assume that total factor productivity follows a first-order autoregressive (AR1) process in logs. The production function is still $y_{it} = \alpha v_{it} + \omega_{it}$ while the productivity process is $\omega_{it} = \rho\omega_{it-1} + \xi_{it}$. We would like to show the properties of GMM estimator $(\widehat{\alpha}, \widehat{\rho})$ using v_{it-1} and $\widehat{\omega}_{it-1}$ as an instrument for v_{it} and $\widehat{\omega}_{it}$, where productivity is fitted based on a guess for α, because the true level of productivity is unobserved. The estimator is now defined as:

Definition 4 *The GMM estimator is a pair* $(\widehat{\alpha}, \widehat{\rho})$ *such that* $\mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] = 0$ *and* $\mathbb{E}\left[\widehat{\xi}_{it}\widehat{\omega}_{it-1}\right] = 0$, *where* $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha})v_{it} + \omega_{it}$ *and* $\widehat{\xi}_{it} = \widehat{\omega}_{it} - \widehat{\rho}\widehat{\omega}_{it-1} = \xi_{it} + (\alpha - \widehat{\alpha})(v_{it} - \rho v_{it-1}) + (\rho - \widehat{\rho})\omega_{it-1} + (\rho - \widehat{\rho})(\alpha - \widehat{\alpha})v_{it-1}$.

In practice, this estimator is solved for iteratively. Because the fitted productivity $\widehat{\omega}_{it}$ depends on α, the econometrician iterates over potential output elasticities $\widehat{\alpha}$ until both the moment conditions for productivity and for the variable input are satisfied. This is why De Loecker and Warzynski (2012) label this procedure "iterative GMM". The estimator, $(\widehat{\alpha}, \widehat{\rho})$, is characterized by the following system of equations defined by the moment conditions:

$$\begin{array}{l} \mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] = 0 \\ \mathbb{E}\left[\widehat{\xi}_{it}\widehat{\omega}_{it-1}\right] = 0 \end{array} \iff \begin{array}{l} (\alpha - \widehat{\alpha})\mathbb{E}\left[(v_{it} - \rho v_{it-1})v_{it-1}\right] + (\rho - \widehat{\rho})\mathbb{E}\left[\omega_{it-1} v_{it-1}\right] + (\alpha - \widehat{\alpha})(\rho - \widehat{\rho})\mathbb{E}\left[v_{it-1}^2\right] = 0 \\ (\alpha - \widehat{\alpha})\mathbb{E}\left[(v_{it} - \rho v_{it-1})\omega_{it-1}\right] + (\rho - \widehat{\rho})\mathbb{E}\left[\omega_{it-1}^2\right] + (\alpha - \widehat{\alpha})(\rho - \widehat{\rho})\mathbb{E}\left[v_{it-1}\omega_{it-1}\right] = 0 \end{array}$$

In general, the above system of equations admits two solutions. One is the true solution with $\widehat{\alpha} = \alpha$ and $\widehat{\rho} = \rho$, while the other solution converges to (α, ρ) as variation in the data increases. We leave the full and formal discussion of this case in Appendix A.2. To understand the essence of the argument consider the following proof sketch, when $\widehat{\alpha}$ and $\widehat{\rho}$ are not too far from α and ρ respectively, the terms of the form $(\widehat{\alpha} - \alpha)(\widehat{\rho} - \rho)$, are of second order. In this case, the system characterizing the estimator $(\widehat{\alpha}, \widehat{\rho})$ reduced locally to the matrix equation $V(B - \widehat{B}) = 0$ where

$$B - \widehat{B} = \begin{pmatrix} \alpha - \widehat{\alpha} \\ \rho - \widehat{\rho} \end{pmatrix} \text{ and } V = \begin{pmatrix} \mathbb{E}[(v_{it} - \rho v_{it-1})v_{it-1}] & \mathbb{E}[\omega_{it-1} v_{it-1}] \\ \mathbb{E}[(v_{it} - \rho v_{it-1})\omega_{it-1}] & \mathbb{E}[\omega_{it-1}^2] \end{pmatrix}.$$

As long as the determinant of V is not zero, the GMM estimator is locally identified and asymptotically unbiased. In Appendix A.2, we show that the GMM estimator is globally identified and asymptotically unbiased as long as there is enough variation in the data.

2.3 Measurement Error and the Two-Stage Procedure

So far, we have abstracted from the first stage of the two-stage GMM procedure by assuming that output is observed without error. In this section we add measurement error to observed output. We first show that our previous results on the consistency of the GMM estimator are unaffected by the presence of measurement error although it reduces the precision of the estimates. We then discuss how a first stage "proxy regression" can remove the measurement error and therefore improve precision of the production function parameter and markup estimates.

As in the baseline framework, assume that firms produce y_{it} using the single variable input v_{it} while being subject to idiosyncratic productivity shocks ω_{it}. Furthermore, assume that the firms' output is observed subject to measurement error, or equivalently, that firms are subject to unexpected productivity shocks that occur after input v_{it} is set. The measurement error is log-additive and denoted by η_{it}. All firms produce along:

$$y_{it} = \alpha v_{it} + \omega_{it} + \eta_{it},$$

where y_{it} denotes observed output or output inclusive of the unexpected productivity shock. We assume that measurement errors at time t are independent of past value of the variable input, that is, $\mathbb{E}\left[\widehat{\eta}_{it} v_{it-1}\right] = 0$. If the econometrician ignores the presence of these measurement errors, the GMM estimator is defined as follows:

Definition 5 *The GMM estimator is* $\widehat{\alpha} \in \mathbb{R}$ *such that the moment* $\mathbb{E}\left[(\widehat{\omega_{it} + \eta_{it}}) v_{it-1}\right]$ *is equal to zero where* $\widehat{\omega_{it} + \eta_{it}} = y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha}) v_{it} + \omega_{it} + \eta_{it}$.

The GMM estimator is characterized by:

$$\mathbb{E}\left[(\widehat{\omega_{it} + \eta_{it}}) v_{it-1}\right] = (\alpha - \widehat{\alpha}) \mathbb{E}[v_{it} v_{it-1}] = 0,$$

where we use the fact that $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right] = 0$. As was the case in the baseline framework, the GMM estimator $\widehat{\alpha}$ of the variable input's output elasticity is equal to α as long as $\mathbb{E}[v_{it} v_{it-1}] \neq 0$. The estimator remains unbiased and unidentified as the additional measurement error only increases the variance of the composite error term $\omega_{it} + \eta_{it}$ in the production function. This point is known and has been discussed, for example, in Blundell and Bond (2000).

If the single-stage GMM estimator is consistent, then why bother purging output from measurement error? There are two advantages to purging. The first is that the increase in the variance of the composite error term $\omega_{it} + \eta_{it}$ in the production function raises the standard errors of the production function estimation. Indeed, a similar derivation to the one in Appendix A.1 yields that the asymptotic variance of estimator is

$$\mathbb{V}ar\left[\widehat{\alpha}\right] \sim \frac{\mathbb{E}\left[v_{it-1}^2\right]}{n \mathbb{E}[v_{it} v_{it-1}]^2} \left(\mathbb{E}\left[\omega_{it}^2\right] + \mathbb{E}\left[\eta_{it}^2\right]\right),$$

which increases in measurement error variance. The second advantage is that purging allows the econometrician to identify true productivity ω_{it}, which is relevant in many applications.[12]

2.3.1 Two-Stage Estimator with an Unbiased First-Stage

To purge output of measurement error, we perform a first-stage proxy regression regression before applying the GMM estimator in the second stage. This is the standard two-stage GMM procedure which has been developed and extensively studied by (Ackerberg et al., 2007, 2015) and frequently applied to markup estimation after De Loecker and Warzynski (2012). We first discuss the properties of the two-stage estimator when a researcher has an unbiased estimate of the measurement errors η_{it}. We then discuss how such an unbiased estimate of the measurement errors can be obtained.

For unbiased estimates $\widehat{\eta}_{it}$ of the measurement error η_{it}, define the GMM estimator as follows:

Definition 6 *The GMM estimator is $\widehat{\alpha} \in \mathbb{R}$ such that the moment $\mathbb{E}\left[\widehat{\omega}_{it} v_{it-1}\right]$ is equal to zero where $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} - \widehat{\eta}_{it} = (\alpha - \widehat{\alpha}) v_{it} + \omega_{it} + \eta_{it} - \widehat{\eta}_{it}$.*

Solving for the above defined GMM estimator where we use $\mathbb{E}\left[\omega_{it} v_{it-1}\right] = 0$ yields $\widehat{\alpha} = \alpha + \frac{\mathbb{E}[(\eta_{it} - \widehat{\eta}_{it}) v_{it-1}]}{\mathbb{E}[v_{it} v_{it-1}]}$. It follows that as long as $\eta_{it} - \widehat{\eta}_{it}$ converges to zero, that is $\widehat{\eta}_{it}$ is asymptotically unbiased and orthogonal to past values of the variable input, v_{it-1}, then the GMM estimator is identified and asymptotically unbiased such that $\widehat{\alpha} = \alpha$. The variance of this estimator is such that $\mathbb{V}ar\left[\widehat{\alpha}\right] \sim \frac{\mathbb{E}\left[v_{it-1}^2\right]}{n\mathbb{E}[v_{it} v_{it-1}]^2} \left(\mathbb{E}\left[\omega_{it}^2\right] + \mathbb{E}\left[(\eta_{it} - \widehat{\eta}_{it})^2\right]\right)$. It follows that if the first-stage estimation $\widehat{\eta}_{it}$ of η_{it} is good, meaning that $\mathbb{E}\left[(\eta_{it} - \widehat{\eta}_{it})^2\right]$ is smaller than $\mathbb{E}\left[\eta_{it}^2\right]$, then the two-stage estimator has potential to improve the precision of the estimates of α. A good first-stage estimation can reduce the variance of the GMM estimator.

How does one perform a good first stage? The aim is to identify measurement errors η_{it} and the identification challenge is to separate the measurement error from the firm's productivity ω_{it}, both of which are unobserved. The difference between the terms is that ω_{it} is observed by the firm, and affects its choice of the variable input. We obtain unbiased estimates $\widehat{\eta}_{it}$ of these errors by running a so-called "proxy regression". The idea is that the demand for material inputs is an invertible function in productivity ($\partial v(\cdot)/\partial \omega_{it} > 0$) such that productivity can be expressed as $\omega_{it} = v^{-1}(v_{it}, \Xi_{it})$, where Ξ_{it} is a set of control variables that determines the firms' demand for the variable input besides productivity ω_{it}. Output can therefore be written as:

$$y_{it} = \alpha v_{it} + v^{-1}(v_{it}, \Xi_{it}) + \eta_{it}.$$

[12] A further benefit of purging output from measurement error is that it allows more sophisticated persistent productivity processes than the linear AR(1).

We can estimate η_{it} by running a non-parametric regression along:

$$y_{it} = \Phi(v_{it}, \Xi_{it}) + \eta_{it},$$

which is the first stage of the two-stage GMM procedure.

The first stage gives unbiased estimates of the measurement errors η_{it} as long as the equation is correctly specified. This means that we must include all relevant variables Ξ_{it} to control for productivity. This is straightforward under perfect competition but requires extra care under imperfect competition. Recall that a cost minimizing firm in our simple framework demands variable inputs along (1). This first-order-condition allows to solve for the productivity as a function of the variable input usage, marginal cost and variable input price:

$$\omega_{it} = (1-\alpha)v_{it} - mc_{it} + w_t.$$

We can substitute this expression for productivity in the production function and use that marginal costs can be written as $mc_{it} = p_{it} - \log \mu_{it}$[13] to get

$$y_{it} = v_{it} - p_{it} + \log \mu_{it} + w_t + \eta_{it}.$$

Under perfect competition, firms are price takers which implies that markups equal to one and prices are orthogonal to firms' choices. The production function, after substituting the expression for productivity, reduces to $y_{it} = v_{it} + w_t + \eta_{it} - p_{it}$. The last two terms are orthogonal to input usage, v_{it}, and input price p_{it}. We then regress output on input usage and time fixed effect to get the predicted residual $\widehat{\eta_{it} - p_{it}}$. The residual of this regression is used in the second stage GMM estimator of Definition 6 which boils down to the moment condition

$$\mathbb{E}\left[\left((\alpha - \widehat{\alpha})v_{it} + \omega_{it} + \eta_{it} - \widehat{\eta}_{it} + p_{it} - \widehat{p}_{it}\right) v_{it-1}\right].$$

Since $\mathbb{E}\left[(\eta_{it} - \widehat{\eta}_{it} + p_{it} - \widehat{p}_{it})v_{it-1}\right] = 0$, thanks to the price taker assumption and the measurement error properties, this estimator is also identified, that is $\widehat{\alpha} = \alpha$.

Under imperfect competition, we can construct an estimate $\widehat{\eta}_{it}$ of the measurement errors η_{it} by using the equation $y_{it} = v_{it} - p_{it} + \log \mu_{it} + w_t + \eta_{it}$. To be precise, we construct $\widehat{\eta}_{it}$ as the residual of the regression of quantity, y_{it}, on the variable input usage v_{it}, on the output price p_{it}, on controls for the markup μ_{it}, and, on time fixed effect w_t.[14] As pointed out by Doraszelski and Jaumandreu (2020), our main objective is to estimate firm-level markup, the needs to control

[13]Note that the expression of the marginal cost $MC_{it} = P_{it}/\mu_{it}$ in log deviation from its mean mc_{it} is equal to $p_{it} - \log \mu_{it}$ up to a constant $\mathbb{E}\left[\log \mu_{it}\right]$. In theory, in the first-stage estimation, we also need to add a constant which is in practice absorbed by the time fixed effect.

[14]In the more general multi-inputs non Cobb-Douglas case, the first-order-condition of the cost minimization problem is not linear in inputs usage and cannot be inverted analytically. Although the functional relationship between productivity and inputs, price and markup is well defined and is approximated by a polynomial of inputs.

for markups in this first stage estimation seems to be contradictory. However, we only need to know that there is a structural relationship between markup and controls, we do not need to know the parameters that governs this relationship. In our simulations and empirical sections, we assume that markups are determined by a firm's market share. We therefore include prices and market shares as controls in the first-stage of our baseline two-stage estimator.

2.3.2 Two-Stage Estimator with a Biased First-Stage

As we saw in the previous section, under imperfect competition, an unbiased first-stage requires the use of price p_{it} and markups controls, such as market share s_{it}, which are not always available to the econometrician. In this section, we discuss the properties of the GMM two-stage estimator when price and markup controls are not included in the first-stage.

When these controls are not included in the first-stage, the GMM estimator is biased. Indeed, under imperfect competition, p_{it} and $\log \mu_{it}$ are not orthogonal to v_{it} which implies that the residual of the regression of output on inputs $\widehat{\eta}_{it}$ is biased and does not asymptotically converge to the true η_{it}. The second stage estimator of Definition 6 is $\widehat{\alpha} = \alpha + \frac{\mathbb{E}[(\eta_{it}-\widehat{\eta}_{it})v_{it-1}]}{\mathbb{E}[v_{it}v_{it-1}]}$ where the second term is not asymptotically zero. In other words, the two-stage estimator is biased.

2.4 Taking Stock

To summarize, the GMM estimator of the production function is asymptotically unbiased when information on quantity is available and the lag of the variable input is used as an instrument. This also holds true when quantity is measured with error. To improve the precision of the GMM estimator, a first stage can be added, in which output is purged of measurement error. Under perfect competition, this first stage does not require price or controls for markups. Under imperfect competition, the first-stage does require a control for price and controls for markups. Finally, if price or markup controls are not available under imperfect competition, the first-stage will create a bias that contaminates the second-stage GMM estimator.

3 Potential Biases

Now that we have established the conditions under which the (two-stage) iterative GMM procedure gives unbiased estimates of the production function and markups, we investigate what happens under misspecification. We analyze the theoretical properties of production function estimates under two common data limitations: (i) the use of revenue data in place of quantity data, and, (ii) misspecification in the production function.

3.1 Revenue versus Quantity

Most firm-level datasets do not contain information on prices and quantities. Researchers usually rely on revenue data to measure a firm's output. In this section we study the bias that this causes and its consequences for the markup computation. The revenue for a firm i at time t, is denoted R_{it} whose log mean deviation is denoted r_{it} and is equal to the sum of the log mean deviation of price, p_{it}, and quantity, y_{it}. Substituting the baseline production function yields:

$$r_{it} = y_{it} + p_{it} = \alpha v_{it} + \omega_{it} + p_{it},$$

where α is the parameters of interest. Let us define the GMM estimator using v_{it-1} as an instrument for v_{it} when revenue is used in place of quantity.

Definition 7 *(GMM estimator on revenue) the GMM estimator is $\widehat{\alpha} \in \mathbb{R}$ such that the moment $\mathbb{E}[\widehat{tfpr}_{it} v_{it-1}]$ is equal to zero where $\widehat{tfpr}_{it} = p_{it} + y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha}) v_{it} + p_{it} + \omega_{it}$.*

In the above definition, $\widehat{tfpr}_{it}$ is an estimate of (log) revenue TFP; the product of price and actual TFP. In our baseline framework, there are no reason for this revenue TFP to be orthogonal to past value of the variable inputs since price can correlate with the variable input. The revenue GMM estimator could therefore be biased. To check the consistency of this estimator, let us solve for $\widehat{\alpha}$ such that $\mathbb{E}[\widehat{tfpr}_{it} v_{it-1}] = (\alpha - \widehat{\alpha}) \mathbb{E}[v_{it} v_{it-1}] + \mathbb{E}[p_{it} v_{it-1}] = 0$ where we use the fact that $\mathbb{E}[\omega_{it} v_{it-1}] = 0$. If $\mathbb{E}[v_{it} v_{it-1}] \neq 0$, we find the following unique value of the estimator:

$$\widehat{\alpha} = \alpha + \frac{\mathbb{E}[p_{it} v_{it-1}]}{\mathbb{E}[v_{it} v_{it-1}]}. \tag{2}$$

This estimator does not have to equal the true value α, which means that the estimator is asymptotically biased if $\mathbb{E}[p_{it} v_{it-1}]$ differs from zero. The bias is determined by the correlation between output price and past value of the variable input, that is the instrument. In a practitioner's setting, this correlation is the sample correlation between price and the instrument. The revenue-based estimates of α can therefore be smaller, larger or equal to the true output elasticity. The same holds for revenue-based markup estimates.

Revenue Markups and Oligopolies What is $\mathbb{E}[p_{it} v_{it-1}]$ in practice? There are no model-free constraints on either its size or sign. If firms face persistent aggregate demand shocks and decreasing returns to scale, for example, positive shocks drive up marginal costs and prices, causing a positive correlation between prices and lagged variable inputs. Conversely, firms that face downward-sloping demand curves must reduce prices to sell additional output, causing a negative correlation. The correlation is zero in case there are no aggregate shocks and firms are atomistic price takers, such that their individual decisions (including input choice) do not affect the equilibrium price. But firms are usually assumed to be price setters when markups are estimated. So what is the bias caused by revenue data under oligopolistic competition?

To answer this question, we need to add more structure on the demand side of our benchmark framework. We want to keep assumptions to the minimum while offering a clear exposition. To this end, we assume a very general invertible demand system, where quantity of all firms depends on price of all firms. We abstract from aggregate shocks that change the price-quantity relationship across periods. For all firms i at time t, the quantity vector $Y = \{Y_{it}\}$ is a function of the price vector $P = \{P_{it}\}$ such that $Y = D_t(P)$. A log-linear approximation yields

$$p_{it} = -\sum_j d_{ijt} y_{jt}, \tag{3}$$

where d_{ijt} is the cross-elasticity of firm i's price with respect to firm j's quantity. In appendix B.1, we derive this approximation formally.[15] With this demand system, we can write (2) as:

$$\widehat{\alpha} = \alpha \left(1 - \sum_j \frac{\mathbb{E}\left[d_{ijt}(v_{jt} + \frac{\omega_{jt}}{\alpha})v_{it-1}\right]}{\mathbb{E}\left[v_{it}v_{it-1}\right]}\right).$$

It follows that the bias due to the use of revenue data is equal to one minus the weighted average of demand elasticities and cross-elasticities among the firms sharing the same production function. This biased estimate of the production function can be used to estimate a firm-level markup based on revenue data, $\widehat{\mu}^R_{it} = \widehat{\alpha}\frac{P_{it}Y_{it}}{W_t V_{it}}$. It follows that the revenue-based markup equals:

$$\widehat{\mu}^R_{it} = \mu_{it} \left(1 - \sum_j \frac{\mathbb{E}\left[d_{ijt}(v_{jt} + \frac{\omega_{jt}}{\alpha})v_{it-1}\right]}{\mathbb{E}\left[v_{it}v_{it-1}\right]}\right),$$

where we use the Hall (1988)'s formula to substitute for the true value of the markup, $\mu_{it} = \alpha\frac{P_{it}Y_{it}}{W_t V_{it}}$. The revenue-based markup is equal to the true markup, up to a constant. It follows that revenue-based markups correlate with true markups while their level might differ.

To understand this point, assume that demand is such that $p_{it} = -\gamma y_{it}$ and that firms are in monopolistic competition. This assumption is satisfied by CES preferences with atomistic firms. In terms of the demand system introduced by equation (3), this assumption implies that $\forall i, d_{iit} = \gamma$ and for $j \neq i$, $d_{ijt} = 0$. Under these assumptions, as pointed out by Bond et al. (2020), the revenue estimator equals the revenue elasticity with respect to the variable input $\widehat{\alpha} = \alpha(1-\gamma) = \frac{\partial y_{it}}{\partial v_{it}}(1 + \frac{\partial p_{it}}{\partial y_{it}}) = \frac{\partial r_{it}}{\partial v_{it}}$. The revenue elasticity and the true markup are identical across firms where the latter is equal to $(1-\gamma)^{-1}$. It follows that the revenue markup is equal to one $\widehat{\mu}^R_{it} = (1-\gamma)^{-1}(1-\gamma) = 1$. When markups are identical across firms sharing the same production function, revenue markups do not contain any information on the true markup.

If firms do not have homogeneous markups, however, estimates of the markup from revenue

[15] We implicitly assume a static demand system, where only current period quantity is affecting current prices.

data will generally not equal one. This is because in models with heterogeneous markups (e.g. Atkeson and Burstein (2008), Kimball (1995) or Klenow and Willis (2016)), demand elasticities differ across firms, while we estimate a single output elasticity of the variable input α. To see this formally, assume that demand is such that $d_{iit} \neq d_{jjt}$ and that firms play a static oligopolistic competition game.[16] In that case, the true markup of firm i is $\mu_{it} = (1 - d_{iit})^{-1}$ and is heterogeneous across firms. Firm's i elasticity of revenue with respect to the variable input is equal to the output elasticity times one minus the demand elasticity:

$$\frac{\partial r_{it}}{\partial v_{it}} = \frac{\partial y_{it}}{\partial v_{it}}(1 + \frac{\partial p_{it}}{\partial y_{it}}) = \alpha(1 - d_{iit}).$$

The estimator on revenue is equal to the output elasticity times one minus the *average* demand elasticities among firms in the sample:

$$\widehat{\alpha} = \mathbb{E}\left[\alpha\left(1 - d_{iit}\frac{v_{it}v_{it-1}}{\mathbb{E}\left[v_{it}v_{it-1}\right]}\right)\right].$$

Computing the revenue markup from this estimate gives $\widehat{\mu}_{it}^R = \mathbb{E}\left[\left(1 - d_{iit}\frac{v_{it}v_{it-1}}{\mathbb{E}[v_{it}v_{it-1}]}\right)\right](1 - d_{iit})^{-1}$, which will not equal one for at least one i. The revenue markup contains information as it correlates with the true value of the markup. The key intuition comes from the fact that when the production function is estimated among a set of firms, the estimated elasticity is equal to an *average* revenue elasticity which in general, is not the same for all firms in the sample. Variation in the revenue markup does reflect the variation in true markup.

What are the moments of these revenue markups under oligopolistic competition? The average revenue markup is equal to the average true markup up to a constant, which is some weighted average of demand elasticities among firms sharing the same production function:

$$\mathbb{E}\left[\log \widehat{\mu}_{it}^R\right] = \mathbb{E}\left[\log \mu_{it}\right] + \log\left(1 - \sum_j \frac{\mathbb{E}\left[d_{ijt}(v_{jt} + \frac{\omega_{jt}}{\alpha})v_{it-1}\right]}{\mathbb{E}\left[v_{it}v_{it-1}\right]}\right).$$

As we discussed above, the (average) log-markup equals zero in case firms have homogeneous markups and price-elasticities of demand. In the case where price elasticities are heterogeneous, the average revenue markup is

$$\mathbb{E}\left[\log \widehat{\mu}_{it}^R\right] = -\mathbb{E}\left[\log(1 - d_{iit})\right] + \log\left(1 - \mathbb{E}\left[d_{iit}\frac{v_{it}v_{it-1}}{\mathbb{E}\left[v_{it}v_{it-1}\right]}\right]\right)$$

which is equal to zero up to a Jensen's inequality.[17] In both cases, the average revenue markup

[16] Here we assumed that $\mathbb{E}\left[d_{iit}\omega_{it}v_{it-1}\right] = 0$. For example this assumption is satisfied when conditional on the past value of the variable input v_{it-1}, productivity ω_{it} and demand elasticity d_{it} are orthogonal. This assumption is also satisfied when conditional on d_{it} productivity ω_{it} and past value of the variable input v_{it-1} are orthogonal. This assumption is in place just to clarify the argument.

[17] Here we are making the similar assumption as in the footnote 16.

has none to little information about the average true markup.[18]

Next, consider the variance of the revenue-based markup and its correlation with the true markup. Since the revenue markup is equal to the true markup up to a constant, after taking log, their variance is equal: $\mathbb{V}ar[\log \widehat{\mu}^{R}_{it}] = \mathbb{V}ar[\log \mu_{it}]$. For the same reason, the correlation between the (log) revenue markup and the (log) true markup is equal to one. While the average revenue markup has close to no information on the true markup, the variance and the dispersion of the (log) revenue markup is informative of the distribution of true markup.

Revenue Markup with Translog Production The result that revenue and quantity markups perfectly correlate depends on the Cobb-Douglas assumption that output-elasticity is a constant. In Appendix B.2 we discuss what happens if the output elasticity is not constant, in the more general case of a translog production function, and show that the main insights remain.

3.2 Cobb Douglas versus Translog

The second bias that we discuss is the one arising from misspecification of the production function. As we have shown in the previous section, the Cobb-Douglas production function has the attractive feature that bias in the estimated level of the markup does not affect the correlation between (log) markup estimates and the true markup. This is why some researchers impose a Cobb-Douglas production function when, for example, data on prices is not available. In this section we analyze the properties of the resulting markup estimates if the true production is more sophisticated. In particular, we assume here that the true production function is translog as in Section 2.2.1, that is, $y_{it} = \alpha v_{it} + \beta v_{it}^2 + \omega_{it}$. The misspecified GMM estimator is:

Definition 8 *(Misspecified GMM estimator) the GMM estimator is $\widehat{\alpha} \in \mathbb{R}$ such that the moment $\mathbb{E}\left[\widehat{\omega}^{Mis}_{it} v_{it-1}\right]$ is equal to zero, where $\widehat{\omega}^{Mis}_{it} = y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha}) v_{it} + \beta v_{it}^2 + \omega_{it}$.*

By wrongly assuming a Cobb-Douglas production function, the econometrician does not take into account that the output elasticity of v_{it} varies with v_{it}. This leads to invalid estimates of productivity. $\widehat{\omega}^{Mis}_{it}$ is the sum of true productivity and an additional term that differs from zero unless $\beta = 0$, even if the estimated $\widehat{\alpha} = \alpha$. As soon as $\beta \neq 0$, that is, when the production function is not log-linear, the productivity estimate at the true value of α is biased.

This biased estimate of the productivity leads to a biased estimate of α. We can solve for the estimator defined above, that is, solve for the $\widehat{\alpha}$ such that $\mathbb{E}\left[\widehat{\omega}^{Mis}_{it} v_{it-1}\right] = (\alpha - \widehat{\alpha})\mathbb{E}[v_{it} v_{it-1}] + \beta \mathbb{E}[v_{it}^2 v_{it-1}] = 0$. If $\mathbb{E}[v_{it} v_{it-1}] \neq 0$, solving for $\widehat{\alpha}$ gives:

[18] Note that in the latter case of heterogeneous demand elasticity, when firms' demand elasticity converge to zero, that is, when firms quantity does not affect their price, markup is converging to one, while the value of the second term is converging to zero. Conversely, when firms' demand elasticity converge to one, the markup diverge to infinity and the second term diverge to minus infinity.

$$\widehat{\alpha} = \alpha + \beta \frac{\mathbb{E}[v_{it}^2 v_{it-1}]}{\mathbb{E}[v_{it} v_{it-1}]}.$$

The misspecified estimator is identified, as there is a unique value of $\widehat{\alpha}$ that satisfies the above definition, but biased as this value is not equal to α. The bias is due to the omitted variable v_{it}^2, which is correlated with the instrument v_{it-1}. To see how the biased production function estimates contaminate markups, substitute the value of $\widehat{\alpha}$ in the Cobb-Douglas version of the Hall (1988)'s formula $\widehat{\mu}_{it}^{Miss} \equiv \widehat{\alpha} \frac{P_{it} Y_{it}}{W_{it} V_{it}}$. The true markup is such that $\mu_{it} = (\alpha + 2\beta v_{it}) \frac{P_{it} Y_{it}}{W_{it} V_{it}}$, which allows us to write the estimated markup as

$$\widehat{\mu}_{it}^{Miss} = \frac{\alpha + \beta \frac{\mathbb{E}[v_{it}^2 v_{it-1}]}{\mathbb{E}[v_{it} v_{it-1}]}}{\alpha + 2\beta v_{it}} \mu_{it}. \tag{4}$$

The estimated markup is equal to the true markup times a factor decreasing in the input usage v_{it}. A direct implication of this is that estimated and true markup do not correlate perfectly, as long as the true markup is correlated with input usage.

What is the implication of this misspecification for the estimated average markup? We can answer this by taking expectations of (4) in logs:

$$\mathbb{E}\left[\log(\widehat{\mu}_{it}^{Miss})\right] = \mathbb{E}\left[\log \mu_{it}\right] + \log\left(\alpha + \beta \mathbb{E}\left[v_{it} \frac{v_{it} v_{it-1}}{\mathbb{E}[v_{it} v_{it-1}]}\right]\right) - \mathbb{E}\left[\log(\alpha + 2\beta v_{it})\right].$$

The average (log) estimated markup is therefore equal to the average true markup up to a Jensen's inequality. Therefore, the average log markup estimated with a misspecified production function contains information about the true level of markups.

We next analyze the second moment of the misspecified markup distribution. From equation (4), the variance of the misspecified markup is equal to $\mathbb{V}ar\left[\log(\widehat{\mu}_{it}^{Miss})\right] = \mathbb{V}ar\left[\log \mu_{it}\right] + \mathbb{V}ar\left[\log(\alpha + 2\beta v_{it})\right] - 2\mathbb{C}ov\left[\log \mu_{it}, \log(\alpha + 2\beta v_{it})\right]$. The variance of the misspecified markup is different from the true markup variance for two reasons. First, the output elasticity $\log(\alpha + 2\beta v_{it})$ is not constant across firms under translog production function, since $\beta \neq 0$. Second, the correlation between markup and the output elasticity can be non-zero. To see this, let us substitute the Hall (1988)'s formula in the covariance between markup and output elasticity to get $\mathbb{C}ov\left[\log \mu_{it}, \log(\alpha + 2\beta v_{it})\right] = \mathbb{V}ar[\log(\alpha + 2\beta v_{it})] + \mathbb{C}ov\left[\log \frac{P_{it} Y_{it}}{W_t V_{it}}, \log(\alpha + 2\beta v_{it})\right]$. The first term is positive, while the second term can be positive or negative. Finally, the covariance between misspecified markup and true markup is given by $\mathbb{C}ov\left[\log \widehat{\mu}_{it}^{Miss}, \log \mu_{it}\right] = \mathbb{V}ar\left[\log \mu_{it}\right] - \mathbb{C}ov\left[\log(\alpha + 2\beta v_{it}), \log \mu_{it}\right]$. This can be positive or negative depending on the covariance between true markups and output elasticity relative to the true markup variance.

To conclude, misspecification in the production function leads to a markup estimate whose average level reflects the average level of true markup, up to Jensen's inequality. However, both correlation and variance of the misspecified markups are different from the true ones.

4 Data

We use administrative data on French manufacturing firms to both quantify our simulations and to empirically analyse the properties of markup estimates. We combine two main datasets. The FARE dataset (*Fichier Approaché des Résultats d'Esane*) provides a detailed balance sheet and income statement while the EAP survey (*Enquête Annuelle de Production*) provides data on both revenues and the quantities of products that firms ship, which we use to obtain a proxy for prices. FARE covers the universe of non-financial French firms and originates from filings to the tax administration (DGFiP). EAP is based on a product-level statistical survey by the statistical office (INSEE) which exhaustively covers manufacturing firms with at least 20 employees or revenue in excess of 5 million euros, and a representative sample of smaller firms.[19]

With the exception of prices, we obtain all variables for the production function estimation from FARE. Revenue is a firm's total sales (including exports),[20] the wage bill (measured as the sum of wages and social security payments), capital (measured by fixed tangible assets on the balance sheet),[21] expenditure on purchased services and expenditure on purchased materials. Materials are defined as physical intermediate goods and raw materials that firms purchase from others. We use NACE Rev. 2 industry codes and define industries j (at which firms have the same production technologies) at the 2-digit level. Market share is defined as the ratio of the firm's revenue over total revenue of all firms in FARE in the 5-digit industry in a given year.

We obtain data on prices from EAP. EAP is a product-level dataset detailing the firms revenue and quantity produced across 10-digit industries.[22] We define a product as the combination of a 10-digit product code and a unit of account.[23] We drop around one-third of firm-products without quantity data. For each combination of a firm and a product we calculate a price as the ratio of revenue over the quantity of the product sold. We then standardize this price by dividing it by the revenue-weighted average price of the product across the entire sample.[24] The firm's price in a year is then given by the sales-weighted average of standardized prices

[19]Smaller firms are sampled with a new sample being drawn annually. Because our production function estimation requires lagged instruments, small firms are not included in the data unless they were randomly sampled for two consecutive years. Our data should therefore be seen as exhaustive of manufacturing firms with at least 20 employees or revenue in excess of 5 million euros.

[20]Data on domestic sales is also available separately, but because we do not have data on the fraction of inputs that account for exports we cannot rely on data on domestic sales to estimate the production function.

[21]We do not rely on the perpetual inventory method because that would require a guess for the firm's initial value of capital. Because our data only covers 11 years, this would lead to a particularly large measurement error (see, e.g., Collard-Wexler and De Loecker 2020). Data on investments is furthermore missing from FARE in 2008. For 2009 to 2018, the correlation between balance sheet capital and estimates of capital from the perpetual inventory method have a correlation of 0.92 to 0.99, depending on the assumed rate of depreciation.

[22]EAP separately reports data for various models of production based on the degree to which the producer is a subcontractor or subcontracts production. We define revenue and quantities for a product as the sum of revenues and quantities over all modes of production for a product in a given firm year.

[23]Examples of units of accounts are kilos, tons or pieces. We combine units of accounts and product codes as firms that use different units of accounts for the same product might produce relatively heterogeneous goods

[24]As a robustness check we standardize prices using the revenue-weighted average price at the 8-digit sector level. The resulting firm-level prices have a 0.89 correlation with prices standardized at the 10-digit product level.

Table 1: Summary Statistics

Variable	Mean	St. Dev.	Median	10th Pct.	90th Pct.	Observations
Revenue	15,853	64,539	27,82	488	28,502	194,702
Quantity	14,154	60,984	1762	210	25,531	194,702
Wage Bill	3,162	12,201	784	176	6,021	194,702
Capital	7,749	34,261	773	98	12,257	194,702
Purchased Materials	7,114	29,154	924	103	12,562	194,702
Purchased Services	3,994	21,608	700	109	6,843	194,702
Normalized Price	9.31	86.37	1.23	.77	6.32	194,702

Note: Summary statistics for French manufacturing firms from 2009 to 2018. Data is obtained from FARE (balance sheet and income statement variables) and EAP (normalized prices). Nominal values are deflated using two-digit EU-KLEMS deflators and are expressed in thousands of 2010 euros. Quantity is measured as firm-deflated revenue. The dataset contains 31,476 unique firms across 206 (18) sectors at the five (two) digit level. All variables are winsorized at their 1% tails.

across the products that it produces. We define quantity as the ratio of revenue over this price.

To deflate input variables we use two-digit industry deflators from EU-KLEMS. This is consistent with the assumption that firms operate on competitive input markets with equal prices across the two-digit level. Revenue is deflated with the gross output deflator, material inputs and purchased services are deflated using the intermediate input deflator. Wages and the capital stock are deflated using the GDP deflator.[25]

We drop firms with missing, zero or negative revenue, material purchases, service purchases, wage bills or capital. We also drop firms in sectors with fewer than 12 firms in a given year to comply with confidentiality requirements.[26] We drop firms without price data in EAP, which restricts the sample to manufacturing firms. We also drop firms with fewer than 2 employees as the number of single-employee firms grows rapidly over our sample due to a regulatory change. To treat for outliers in the remaining sample we winsorize sales, quantity, prices, material and service inputs, and capital at the 1% level within two-digit industries. The resulting sample contains 194,702 firm-years for 31,476 unique firms across 206 five-digit sectors. Summary statistics are provided in Table 1. A description of the two-digit sectors included in our analysis is provided in Table 2.

5 Simulation

In this section we analyse estimates of production function elasticities and markups in cases where the true elasticities and markups are known. To do so, we estimate the production function for a set of simulated firms in a rich macroeconomic model. Firms are heterogeneous in their productivity, the quantity of a fixed input at their disposal, and therefore the market share

[25] At the time of writing, the most recent year for EU-KLEMS deflators is 2017. To deflate 2018 variables we extrapolate the price index using the sector's average inflation in other years.

[26] We calculate market share before restricting the sample.

Table 2: Sectors (two-digit) in the Cleaned EAP-FARE Dataset

Description	NACE code	Observations
Manufacturing of ...		
... textiles	13	7,499
... wearing apparel	14	6,025
... leather and related products	15	2,500
... wood and of products of wood and cork, except furniture	16	10,517
... paper and paper products	17	6,913
... printing and reproduction of recorded media	18	11,275
... chemicals and chemical products	20	9,170
... rubber and plastic products	22	19,251
... other non-metallic mineral products	23	15,519
... basic metals	24	4,816
... fabricated metal products, except machinery and equipment	25	28,818
... computer, electronic and optical products	26	7,170
... electrical equipment	27	8,161
... machinery and equipment n.e.c.	28	18,067
... motor vehicles, trailers and semi-trailers	29	6,002
... other transport equipment	30	1,030
... furniture	31	12,486
Other manufacturing	32	5,787
Repair and installation of machinery and equipment	33	13,696

that they achieve. Heterogeneous market shares cause differences in markups across firms, determined endogenously as a consequence of oligopolistic competition. The model is outlined in Section 5.1 while the calibration is detailed in 5.2. Results are presented in 5.3.

5.1 Model

We analyze a single sector. In line with our setup in Section 2, a sector is defined as a collection of firms that have the same structural parameters of their production function and that face the same prices on input markets.

Demand We choose a market structure that allows firms to have heterogeneous markups that are determined by a combination of structural parameters and their market share. Following Atkeson and Burstein (2008), we implement this by assuming that firms compete in a double-nested CES demand system. The sector consists of a discrete N markets, where a market is defined as a group of firms that compete oligopolistically with one another. Output across markets is aggregated to the sector level along:

$$Y_t = \left[\sum_{h=1}^{N} Y_{ht}^{\frac{\sigma-1}{\sigma}} \right]^{\frac{\sigma}{\sigma-1}}, \tag{5}$$

where h indexes markets and where σ denotes the elasticity of substitution across market-level goods. Output of the market-level good Y_{kt} is the aggregate of firm-level output across the N_h

firms that operate in *h* along:

$$Y_{ht} = \left[\sum_{i=1}^{N_h} Y_{iht}^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}}, \tag{6}$$

where Y_{iht} denotes the output of firm *i* and where ε denotes the elasticity of substitution across firm-level goods within a market. Following Atkeson and Burstein (2008) we assume that $\varepsilon > \sigma$, reflecting that it is easier to substitute goods across firms than across markets. The double-nested CES system gives rise to the standard demand function for firm *i*'s output:

$$Y_{iht} = \left(\frac{P_{iht}}{P_{ht}} \right)^{-\varepsilon} Y_{ht} \quad \text{where} \quad Y_{ht} = \left(\frac{P_{ht}}{P_t} \right)^{-\sigma} Y_t, \tag{7}$$

where aggregate demand $P_t^{\sigma} Y_t$ is exogenous and where P_{ht} is the usual CES market price index:

$$P_{ht} = \left(\sum_{i=1}^{N_h} P_{iht}^{1-\varepsilon} \right)^{\frac{1}{1-\varepsilon}}. \tag{8}$$

The price-setting firm internalizes that P_{ht} increases when it raises its own prices. Following Burstein et al. (2020), we assume for tractibility that firms do not internalize that it may induce an increase in the overall price level P_t. As such, we assume that firms behave as if markets are atomistic (as in Atkeson and Burstein 2008), despite the actual setup featuring a finite number of markets. Under Cournot competition, the resulting profit-maximizing markup reads:[27]

$$\mu_{iht} = \frac{\varepsilon}{\varepsilon - 1} \left(1 - \frac{\frac{\varepsilon}{\sigma} - 1}{\varepsilon - 1} s_{iht} \right)^{-1}, \tag{9}$$

where market share s_{iht} is defined as the firm's share in market revenue:

$$s_{iht} = \frac{P_{iht} Y_{iht}}{P_{h(i)t} Y_{h(i)t}}, \tag{10}$$

where subscript $h(i)$ indicates the market in which firm *i* operates. The firm's markup ranges from $\varepsilon/(\varepsilon - 1)$ for a firm whose market share approaches zero to $\sigma/(\sigma - 1)$ for a monopolist, which is higher than the small firm's markup given the assumption that $\varepsilon > \sigma$.

Technology Firms produces their a variable input V_{iht} and a fixed input K_{iht}, with log-inputs respectively denoted by v_{iht} and k_{iht}. The production function for log output y_{iht} is translog:

$$y_{iht} = \omega_{it} + \gamma \alpha v_{iht} + \gamma (1 - \alpha) k_{iht} + \gamma \frac{\alpha(1-\alpha)}{2} \frac{\phi - 1}{\phi} \left(v_{iht}^2 + k_{iht}^2 - 2 k_{iht} v_{iht} \right), \tag{11}$$

[27] Bertrand competition, where the firm's first order condition takes prices rather than quantities as given, yields a similar expression. See Atkeson and Burstein (2008), Grassi (2018), and Burstein et al. (2020) for an elaborate discussion. For the purpose of these simulations we require that markups are determined by demand elasticities and market share, which is it the case for both Cournot and Bertrand competition.

where ω_{it} is the log of (hicks-neutral) total factor productivity, γ measures the degree of returns to scale, α determines the weight of the variable input in the production function, while ϕ approximates the elasticity of substitution between the flexible and the fixed input. Our log production function (11) is motivated by the following generalized constant elasticity of substitution production function:

$$Y_{iht} = \Omega_{iht} \left(\alpha V_{iht}^{\frac{\phi-1}{\phi}} + (1-\alpha) K_{iht}^{\frac{\phi-1}{\phi}} \right)^{\frac{\phi}{\phi-1}\gamma},$$

where $\Omega_{iht} \equiv \exp \omega_{iht}$. In Appendix B we show that this production function converges to the Cobb-Douglas production function as $\phi \to 1$ and that an approximation of the production function around $\phi = 1$ yields translog function (11). We specify the production function with a constant degree of homogeneity (γ) such that the model admits an analytical expression for (size-dependent) marginal costs, facilitating the calculation of the equilibrium.

Equilibrium We consider the following partial equilibrium. Given an exogenous sequence for variable input prices W_t, aggregate demand $P_t^\sigma Y_t$, productivities ω_{iht} and fixed factors k_{iht}, the equilibrium is defined as a sequence of markups μ_{iht}, prices P_{iht}, log marginal costs mc_{iht}, market shares s_{iht}, log variable inputs v_{iht},and log outputs y_{iht} and sector price indices P_{ht} such that output follows from demand function (7), sector price indices follow from (8), markups are set along equation (9), market share is given by (10), prices are the product of marginal costs and markups, variable inputs are in line with the firm's first order condition and marginal costs follow from (11), for all i, *h* and *t*. Derivations are provided in Appendix C.

5.2 Calibration

We simulate the behavior of 1440 firms, which is the average number of firms in two-sector industries in the EAP data. We divide these firms into 180 markets and simulate the economy for 40 periods.[28] There are 13 parameters, each of which we calibrate externally. The parameters are summarized in Table 3. In calibrating the model, we are constrained by the fact that the true values of many parameters (such as those of the production function and the productivity process) are in fact the object of our empirical analysis. Our approach is therefore to assume reasonable values in line with the literature as an example quantification.

There are two aggregate shocks: aggregate demand $P_t^\sigma Y_t$ and prices of the variable input W_t. We assume both series follow a log-linear first-order autoregressive process with persistence ρ_{PY} and ρ_W, respectively, with shocks $\xi_{PY} \sim N(0, \sigma_{PY})$ and $\xi_W \sim N(0, \sigma_W)$. Fluctuations in ag-

[28]Recall that markets define the level at which firms compete on product markets. By modeling a large number of small markets rather than a small number of large numbers we reduce the computational complexity of the simulation. An appropriate calibration of the volatility of the productivity process assures that markets have realistic levels of concentration and sufficient markup dispersion.

Table 3: Overview of Example Parameter Calibration for Simulation

Parameters	Value	Description
α	0.4	Share of variable input
γ	0.8	Returns to scale
ϕ	1.1	Elasticity of substitution
σ	1.1	Demand elasticity across markets
ε	10	Demand elasticity across firms in a market
N, N_h	180, 8	Number of markets and firms per market
ρ^w, σ^w	0.87 , 0.06	AR(1) persistence and std. dev. of W_t
ρ^D, σ^D	0.78 , 0.19	AR(1) persistence and std. dev. of $P_t^{-\sigma}Y_t$
ρ^ω, σ^ω	0.70 , 0.10	AR(1) persistence and std of firm-level ω_{it}
ρ^k, σ^k	0.66 , 0.66	AR(1) persistence and std of firm-level k_{it}
$\tilde{\sigma}^\eta$	0.122	std. dev. meas. error on output

Note: a detailed description of the calibration is provided in the text. EAP-FARE refers to the cleaned dataset of French manufacturing firms, exhaustive for firms with at least 20 employees or 5 million euros in revenue, for 2009-2018. EU-KLEMS IIP refers to the two-digit sector intermediate input price index while EU-KLEMS VA refers to detrended nominal two-digit value added.

gregate demand assure that the relationship between output and market share vary over time. Fluctuations in the price of the variable input assures that firms' lagged productivity and lagged variable inputs are not co-linear after conditioning on the fixed inputs, which is needed to be able to separately identify the productivity process and the production function parameters, as discussed in Section 2. To calibrate the process for the price of the flexible inputs, we estimate an autoregressive process for the price index of intermediate inputs from sector-level manufacturing data in EU-KLEMS. We run simple autoregressive regressions for the log of the index and find an autoregressive coefficient ρ^w of 0.87 at the 2-digit sector level when controlling for industry and year fixed effects. Residuals have a standard deviation σ^w of 0.06.[29] For aggregate demand $P_t^{-\sigma}Y_t$ we estimate a similar autoregressive process, using the detrended sector-level nominal value added as the dependent variable.[30] We find a high degree of persistence in aggregate demand with a ρ^D of 0.78 while the residuals have a standard deviation of 0.19.

There are two sources of firm heterogeneity in the model: the firm's log-endowment of the fixed input f_{iht} and the firm's log-total factor productivity ω_{it}. Both evolve exogenously through linear first order autoregressive processes with persistence ρ_f and ρ_ω, respectively, and are subject to shocks $\xi_f \sim N(0, \sigma_f)$ and $\xi_\omega \sim N(0, \sigma_\omega)$. Both sources of firm heterogeneity are similar in that firms with either higher productivities or higher values for the exogenous fixed input have, ceteris paribus, greater output. They are different in that the fixed input is observable, while productivity is not. To calibrate the persistence and volatility of the fixed factor, we run autoregressive regressions on the log of capital for firms in the EAP data. We find a persistence pa-

[29] Appendix Table A2 presents AR(1) coefficients for various specifications, which suggest a narrow range of 0.86 to 0.90 for the AR(1) coefficient and 0.042 to 0.046 for the standard deviation of the shocks.

[30] We detrend $P_t^{-\sigma}Y_t$ using nominal GDP to account both for trend increases in prices and for aggregate growth in order to obtain a stationary nominal series. Results are similar when detrending with the GDP deflator.

rameter ρ^k of 0.66 and a volatility of shocks σ_k of 0.66.[31] A particular challenge is the estimation of the persistence ρ^ω and volatility σ^ω of the productivity process. To obtain these empirically requires knowledge of the parameters of the production function, which is the objective of our analysis. We take the pragmatic approach of calibrating ρ^ω and σ^ω in line with common values of the literature, and assure that these values are in line with our findings in Section 4. We set ρ^ω to 0.6 in line with Decker et al. (2020) and set productivity volatility σ^ω to 0.1.

When calibrating the production function parameters, we think of purchased materials as v_{iht} and a composite of all other factors as the fixed factor k_{iht}. We calibrate the variable input share parameter α to 0.4 to match the average ratio of expenditure on materials over revenue in the EAP-FARE data, which is 0.38. We calibrate the returns to scale parameter γ to 0.8 in order to have a modest degree of decreasing returns to scale, in line with the estimate by Basu and Fernald (1997). We assume an elasticity of substitution ϕ of 1.1 as purchased materials include intermediate inputs from other firms, which can substitute in-house production.

We introduce measurement error in observed quantity y_{iht}, denoted by η_{iht}, after performing the simulations. This is the error that we aim to purge in the first stage of the iterative GMM procedure. We assume that $\eta_{iht} \sim N(0, \sigma_y \tilde{\sigma}_\eta)$, where σ_y is the standard deviation of true output across all firm-years in the sector and $\tilde{\sigma}_\eta$ is a scalar that determines the magnitude of measurement error relative to standard deviation of true output. We calibrate $\tilde{\sigma}_\eta$ to 0.122, in line with the relative variance of output and fitted values of a regression of output on prices, market share, time fixed effects and a third degree polynomial in the firms' inputs in EAP.

5.3 Results

We take the simulated firm-level data on revenue, output and inputs and use it to estimate markups along the two-step iterative GMM procedure. We estimate various alternative specifications of the procedure to match the approaches typically followed in the literature. In all specifications we assume that the researcher correctly assumes that the variable input is v_{iht}. This is therefore also the variable that we use to calculate markups after estimating the production function. We further assume that the researcher chooses the moment condition such that the residual of the AR(1) process for productivity, ξ_{iht}, is orthogonal to the lagged variable input and the contemporaneous fixed input.[32]

We start by estimating a preferred specification in order to establish that it is feasible to estimate the production function parameters and markups in our setup. The preferred specification has three components, each of which we deviate from in subsequent specifications. Firstly, we

[31] Appendix Table A3 presents the AR(1) estimates for capital.

[32] We add a constant to the second stage auto-regressive productivity estimation, so that no constant need to be added to the production function itself. It is straightforward to show that estimating a constant in the production function or estimating a constant in the AR process is equivalent. If both are included in the procedure then they are not identifiable, as one can identify only a linear combination of the two.

estimate our preferred specification using quantity as the measure of a firm's output, hence we assume that the researcher perfectly observes the prices that firms set. Secondly, we estimate the preferred specification using a theoretically valid first-stage regression. As shown in Section 2, to correctly control for ω_{iht} in the regression that purges measurement error η_{iht}, the control variables must account for the log of marginal costs. To do so, we include both the log of the price and the firm's market share as additional controls in the first stage, where the latter is the proxy for the markup.[33] Combined with a third order expansion of the inputs v_{it} and k_{it}, this should allow us to identify the measurement error η_{iht} with reasonable precision. Thirdly, the preferred specification estimates a production function of the translog form, in line with (11):

$$y_{iht} = \beta_v v_{iht} + \beta_k k_{iht} + \beta_{vv} v_{iht}^2 + \beta_{kk} k_{iht}^2 + \beta_{vk} k_{iht} v_{iht} + \omega_{iht}, \tag{12}$$

where the true values of the production functions satisfy the following relations with the true production function parameters (α, γ, η):

$$\beta_v = \gamma\alpha, \quad \beta_k = \gamma(1-\alpha), \quad \beta_{vv} = \gamma \frac{\alpha(1-\alpha)}{2} \frac{\phi - 1}{\phi}, \quad \beta_{kk} = \beta_{vv}, \quad \beta_{vk} = -2\beta_{vv}.$$

Note that we do not impose these theoretical restrictions when estimating (12). We then consider various 'imperfect' specifications of the two-stage GMM procedure and see how the production function and markup estimates change.

Our estimates of the translog production function parameters are presented in Table 4. Coefficients in the column titled 'True' are directly calculated from the deep production function parameters (α, γ, ϕ). The three subsequent columns present estimates of the production function where output is measured in quantities while the final three columns present estimates where revenue is used. Bootstrapped standard errors are in parentheses.

Our preferred specification is presented in the second column, where 'Full' indicates that the first stage includes log price and market share controls. The estimates show that the preferred specification is able to identify the parameters of the production successfully. All coefficients are within one tenth of a decimal point of their true value, which is also the case for the model's deep parameters. Bootstrapped standard errors are generally small and coefficients are highly significant.[34] The markups which arise from these production function estimates are summarized in the second row of Table 5. Results for the preferred specification with a translog production function, quantity data and a full first stage are closely in line with true markups. The estimated markups have a correlation of 1.00 with the true markup, although the level and standard deviation of the markup are estimated with slight error. This is in line with the modest

[33] We do not include a polynomial of market share to control for non-linearities in the markup-market share relationship because the correlation between market share and its square exceeds 0.99.

[34] Standard errors differ from zero because (1) the first stage approximates the implicit relationship between productivity and inputs through a third-order polynomial and (2) market share proxies imperfectly for markups.

Table 4: Estimated Translog production function parameters by Specification

Coefficients	True	Quantity Full	Quantity Basic	Quantity None	Revenue Full	Revenue Basic	Revenue None
$\beta_v = \alpha\gamma$	0.32	0.32	0.32	0.44	0.67	0.67	0.87
		(0.012)	(0.013)	(0.031)	(0.031)	(0.029)	(0.047)
$\beta_k = (1-\alpha)\gamma$	0.48	0.47	0.48	0.42	0.21	0.21	0.12
		(0.007)	(0.007)	(0.015)	(0.022)	(0.02)	(0.027)
$\beta_{vv} = \gamma\frac{\alpha(1-\alpha)}{2}\frac{\phi-1}{\phi}$	0.009	0.009	0.008	0.032	0.034	0.032	0.073
		(0.002)	(0.003)	(0.006)	(0.007)	(0.007)	(0.01)
$\beta_{kk} = \beta_{vv}$	0.009	0.009	0.007	0.012	0.005	0.005	0.011
		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.003)
$\beta_{vk} = -2\beta_{vv}$	-0.017	-0.019	-0.016	-0.037	-0.032	-0.030	-0.064
		(0.003)	(0.003)	(0.006)	(0.008)	(0.008)	(0.01)
Implied parameters:							
α	0.400	0.405	0.403	0.511	0.762	0.757	0.875
γ	0.800	0.798	0.797	0.864	0.880	0.879	0.990
ϕ	1.100	1.110	1.092	1.207	1.251	1.226	2.475

Note: Estimated production function coefficients for different specifications. The top panel presents production function estimates. The bottom panel presents the deep parameters implied by the estimated production function.The first column presents true values for comparison. Bootstrapped standard errors are in parentheses. Full, Basic and None describe the specification of the first-stage regressions. Full first-stages include a third order expansion in the production function inputs, time fixed effects and additional controls for log price and market share. Basic first-stages do not include the additional controls. Columns headed None do not deploy a first stage and therefore estimate markups on output variables that include measurement error η_{iht}.

differences between the true and estimated production function parameters in Table 4.

Revenue versus Quantity We next deviate from the preferred specification by using revenue instead of quantity data to estimate the production function. In the fifth column of Table 4 we report a significantly higher estimate for β_v, which increases from 0.32 to 0.67. Conversely, we find a significantly lower linear coefficient for the fixed input, β_k, falling from 0.47 to 0.21. The increase in β_v might be surprising, because the revenue elasticity of an input should fall short of the quantity elasticity when demand curves are sloping downward. Recall, however, that equation (2) showed that revenue-based coefficients can be biased upwards, downwards or be unaffected, depending on the correlation between prices and inputs over all firms in the sectors. The result that log-markups average zero up to Jensen's inequality was conditional on the absence of time fixed effects. Firms in our simulation are subject to aggregate demand shocks, which create a positive correlation between input usage and prices under diminishing returns to scale. In line with this, we find that log-prices have a 0.42 correlation with log-variable inputs in our simulated data. Controlling for time fixed effects the correlation is negative, as expected.

In the first row of the bottom panel of Table 5 we compare estimates of the markup based on the revenue-production function. The level of the markup is overestimated, which is caused by the

Table 5: Overview - Translog Productivity and Log Markup Estimates

	Correlation ln $\widehat{\mu}_{iht}$ with true markup	Log Markup Moments			
		Mean	St. Dev.	Median	IQR
True values	1.00	0.238	0.042	0.220	0.239
Quantity					
Full first stage (preferred)	1.00	0.268	0.063	0.221	0.342
Basic first stage	1.00	0.265	0.067	0.220	0.366
No first stage	0.45	0.316	0.167	0.204	1.156
Revenue					
Full first stage	0.77	0.799	0.118	0.713	0.222
Basic first stage	0.80	0.798	0.115	0.714	0.213
No first stage	0.38	0.813	0.277	0.651	0.588

Note: Table of moments of estimated productivity. The first column present correlations of estimated markups with the true values. Full first-stages include a third order expansion in the production function inputs and additional controls for log price and market share. Basic first-stages do not include the additional controls.

overestimation of β_v. We do find that the revenue-based markups are highly informative of true markups, with a point correlation of 0.77 between both. We conclude that these results confirm that the revenue-based estimates of the production function elasticities are *not* the revenue elasticities of an input. If they were, Bond et al. (2020) show that the log markups should equal 0 and be uninformative of true markups. Instead, one should think about the revenue-based elasticities as *biased* estimates of output elasticities of the inputs. This bias causes the log-markup to *average* 0 up to a Jensen's inequality, or to have a further bias due to the presence of time effects such as demand shocks.

First Stage In a second deviation from the preferred specification, we compare production function estimates with different first stages. Results so far include a third-order expansion of the inputs as well as additional control variables for price and market share. We next consider a 'basic' first stage where we drop the control variables for price and market share. The resulting first-stage specification is frequently used in markup estimations (e.g. De Loecker and Warzynski 2012 De Loecker et al. 2020). We find similar estimates for the parameters in the production function when using the basic first stage. The linear coefficients β_v and β_k are unaffected for both quantity and revenue-based estimations, although all the higher-order terms are slightly underestimated. Looking at the correlation with the true (log) markup, 5 shows that markups from the basic first stage again have a correlation of 1.00 with true markups. The moments of the markup distribution are similar to the full-first stage markups. Conversely, we find that the accuracy of the production function estimates declines sharply when no first stage is conducted.[35] In this case, the production function estimation becomes a simple IV-GMM esti-

[35]The importance of the first stage increases in the variance of measurement error η_{iht}. If the measurement error is reduced to zero, estimates without a first stage become equivalent to estimates from the full first stage.

Table 6: Estimated Cobb-Douglass production function parameters by Specification

Coefficients	True	Quantity Full	Quantity Basic	Quantity None	Revenue Full	Revenue Basic	Revenue None
$\beta_v = \alpha\gamma$	0.32	0.31	0.31	0.32	0.52	0.52	0.52
		(0.003)	(0.002)	(0.002)	(0.003)	(0.003)	(0.003)
$\beta_k = (1-\alpha)\gamma$	0.48	0.49	0.49	0.49	0.3	0.3	0.29
		(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
Implied parameters:							
α	0.40	0.39	0.39	0.40	0.64	0.64	0.64
γ	0.80	0.81	0.80	0.81	0.81	0.81	0.82
ϕ	1.10	1.00	1.00	1.00	1.00	1.00	1.00

Note: Estimated production function coefficients for different specifications. The top panel presents production function estimates. The bottom panel presents the deep parameters implied by the estimated production function.The first column presents true values for comparison. Bootstrapped standard errors are in parentheses. Full, Basic and None describe the specification of the first-stage regressions. Full first-stages include a third order expansion in the production function inputs, time fixed effects and additional controls for log price and market share. Basic first-stages do not include the additional controls. Columns headed None do not deploy a first stage and therefore estimate markups on output variables that include measurement error η_{iht}.

mation with lagged variable inputs used to instrument for current variable inputs. The unidentified measurement error causes an overestimation of the output elasticity of the variable input. This causes an overestimation of the average log markup in Table 5 and a significant reduction (from 1.00 to 0.45) in the correlation between estimated and true markups.

Cobb-Douglas In a third deviation, we perform the production function estimation on the same data while assuming that production is Cobb-Douglas:

$$y_{iht} = \beta_v v_{iht} + \beta_k k_{iht} + \omega_{iht}, \tag{13}$$

which is akin to forcing the higher-order terms in (12) to equal 0 and the elasticity of substitution between the variable input and the fixed input (ϕ) to be 1. Table 6 presents the estimates. We find that the Cobb-Douglas estimates of the linear coefficients are closely in line with the true values. Looking at the implied deep parameters, the estimations are within a tenth of a decimal for both the variable input share α and for the degree of returns to scale γ. Cobb-Douglas production function estimates from revenue data show a similar bias as translog estimates: the coefficient for the variable input is upwardly biased, while the fixed correlation with the fixed input is negatively biased.

Table 7 presents the moments of the resulting log markups correlates these markups with the true ones.[36] Because all firms have the same variable elasticity, the correlation between log-markup estimates and true markups is equal across specifications at 0.82. The same holds for the dispersion of markups, which is dictated by the dispersion of the ratio of revenue over

[36]A correlation matrix across all markup specifications is provided in Appendix Table A6.

Table 7: Overview - Cobb-Douglas Productivity and Log Markup Estimates

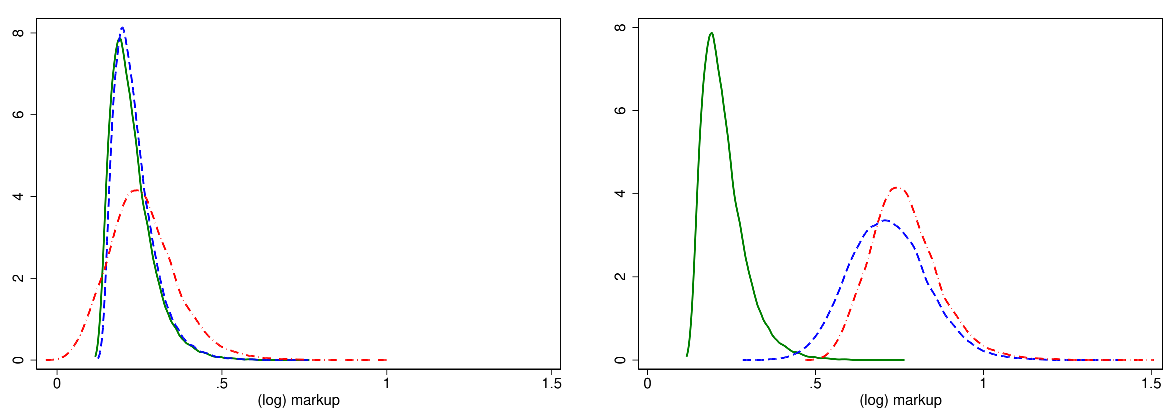
	Correlation $\ln \widehat{\mu}_{iht}$ with true markup	Log Markup Moments			
		Mean	St. Dev.	Median	IQR
True	1.00	0.238	0.042	0.220	0.239
Quantity					
Full first stage	0.82	0.280	0.100	0.250	0.530
Basic first stage	0.82	0.270	0.100	0.250	0.540
No first stage	0.82	0.300	0.100	0.270	0.480
Revenue					
Full first stage	0.82	0.780	0.100	0.760	0.170
Basic first stage	0.82	0.780	0.100	0.760	0.180
No first stage	0.82	0.780	0.100	0.760	0.170

Note: Table of moments of estimated markups. The first column presents correlations markups with the true values. Full first-stages include a third order expansion in the production function inputs and additional controls for log price and market share. Basic first-stages do not include the additional controls.

spending on the variable input. The Cobb-Douglas markup have a standard deviation of 0.1 in log, which is more than twice as high as the standard deviation of true markups.

An illustration of the bias in both the level and the dispersion of the markup arising from assuming a Cobb-Douglas instead of a translog production function is illustrated in Figure 1. It provides kernel plots for the distribution of log markups for the true markups (green-solid), markups from the translog production function (blue-dashed) and the Cobb-Douglas production function (red-dash dotted). The left-hand figure presents results based on revenue, the right-hand side presents results based on quantity. The graph reaffirms the results from the

Figure 1: Distribution of (Log) Markup by Output Variable and Production Function



(a) Quantity (b) Revenue

Notes: The figure plots the distribution of markups (in logs). Green-solid lines present the distribution of the true markups. Blue-dashed lines present markup estimates based on a translog production function. Red-dashed dotted lines present markups based on a Cobb-Douglas production function. The first stage specification includes price and market share controls.

tables. Cobb-Douglas estimates of the markup are similar in terms of average values, but understate the dispersion of the markup. This is because the Cobb-Douglas estimation forces the elasticity of substitution between variable and fixed production factors to equal 1. If the actual elasticity of substitution exceeds 1, firms with higher fixed-factor endowments have lower output elasticities with respect to the variable input, because β_{vk} is negative. This reduces the markup of these firms, which generally have high markups because of their size. Hence this lowers markup dispersion. Conversely, with $\phi < 1$, the coefficient β_{vk} is positive and the overall variable input elasticity is higher for these firms. This raises markup dispersion.

6 Empirics

This section describes the results from the production function and markup estimation on the French EAP-FARE manufacturing data from 2009 to 2018. We start by assessing the elasticities of quantity and revenue with respect to materials. We then compare the levels and dispersion of markups from various specifications, and assess the correlation between the various markup estimates. Finally, we look at how estimated key relationships such as the markup-market share and the markup-profit relationship depend on production function specifications.

6.1 Production Function Estimates

In line with our simulations, we estimate the production function in twelve specifications by running all possible combinations of a Cobb Douglas or translog production function, quantity or revenue as the measure of output, and either the full first stage (with price and market share controls), basic (with only the polynomial in inputs), or absent first stage. We assume that the production function consists of the following (log) inputs: materials m_{it}, the wage bill l_{it}, capital k_{it} and services o_{it}. Following Burstein et al. (2020) we assume that materials involve no adjustment costs and therefore correspond to the freely-set variable input v_{iht} in Section 2.

Table 8 presents the estimated material elasticities θ_{it}^{M} for each of our specifications. Specifications in the first six columns use quantity as the measure of output. Specifications for the final six columns use revenue as the dependent variable. Columns headed *CD* present estimated elasticities from the Cobb-Douglas production function at the two-digit level while columns headed *TL* present average elasticities calculated along

$$\theta_{it}^{M} = \beta_m + 2 \cdot \beta_{mm} m_{it} + \beta_{mo} o_{it} + \beta_{ml} l_{it} + \beta_{mk} k_{it},$$

where the production function elasticities β_x are constant within two-digit sectors but θ_{it}^{M} is heterogeneous across firms within the sector. We report the standard deviation of the firm-level translog elasticities in parentheses.

Table 8: Estimated Material-Output Elasticity by Sector and Specification

	Quantity						Revenue					
	Full		Basic		None		Full		Basic		None	
NACE	CD	TL	CD	TL	CD	TL	CD	TL	CD	TL	CD	TL
Avg.	0.5	0.55	0.46	0.47	0.49	0.56	0.44	0.41	0.39	0.42	0.35	0.41
		(0.43)		(0.30)		(0.55)		(0.19)		(0.19)		(0.18)
13	0.29	0.45	0.18	0.41	0.28	0.46	0.18	0.41	0.18	0.41	0.26	0.42
		(0.23)		(0.21)		(0.26)		(0.16)		(0.16)		(0.18)
14	0.35	0.54	0.21	0.43	0.33	0.54	0.40	0.32	0.40	0.32	0.02	0.36
		(0.28)		(0.25)		(0.32)		(0.24)		(0.24)		(0.37)
15	0.30	0.42	0.26	0.38	0.26	0.38	0.09	0.33	0.09	0.32	0.13	0.37
		(0.16)		(0.19)		(0.20)		(0.23)		(0.21)		(0.35)
16	0.58	0.6	0.55	0.53	0.56	0.57	0.55	0.47	0.55	0.72	0.48	0.46
		(0.13)		(0.13)		(0.14)		(0.11)		(0.32)		(0.18)
17	0.42	0.53	0.44	0.54	0.44	0.53	0.48	0.83	0.47	0.54	0.39	0.46
		(0.17)		(0.18)		(0.13)		(0.34)		(0.21)		(0.13)
18	0.31	0.36	0.31	0.37	0.30	0.35	1.19	0.30	0.05	0.30	0.24	0.29
		(0.16)		(0.17)		(0.15)		(0.11)		(0.11)		(0.11)
20	0.41	0.91	0.31	0.72	0.43	0.88	0.65	0.49	0.64	0.48	0.36	0.51
		(0.56)		(0.42)		(0.58)		(0.17)		(0.16)		(0.19)
22	0.59	0.61	0.55	0.53	0.58	0.61	0.45	0.45	0.45	0.45	0.44	0.45
		(0.20)		(0.18)		(0.20)		(0.13)		(0.13)		(0.13)
23	0.46	0.46	0.41	0.4	0.48	0.45	0.32	0.34	0.32	0.35	0.38	0.44
		(0.07)		(0.07)		(0.10)		(0.11)		(0.12)		(0.15)
24	0.65	0.63	0.77	0.71	0.65	0.63	0.61	0.44	0.61	0.44	0.46	0.43
		(0.23)		(0.17)		(0.25)		(0.19)		(0.19)		(0.21)
25	0.41	0.41	0.42	0.40	0.40	0.40	0.39	0.37	0.39	0.37	0.35	0.36
		(0.21)		(0.17)		(0.19)		(0.16)		(0.16)		(0.17)
26	0.85	0.75	0.57	-0.31	0.81	0.89	0.36	0.40	0.36	0.40	0.39	0.41
		(0.66)		(0.62)		(0.66)		(0.13)		(0.14)		(0.18)
27	0.43	0.6	0.48	0.58	0.42	0.61	0.33	0.46	0.77	0.46	0.38	0.45
		(0.32)		(0.21)		(0.33)		(0.13)		(0.13)		(0.13)
28	0.24	0.21	0.66	0.59	0.32	-0.03	0.31	0.43	0.31	0.43	0.35	0.43
		(0.35)		(0.17)		(0.69)		(0.16)		(0.15)		(0.14)
29	1.07	1.03	0.47	0.76	0.68	1.02	0.54	0.54	0.54	0.54	0.50	0.53
		(0.45)		(0.26)		(0.48)		(0.15)		(0.15)		(0.21)
30	0.24	0.30	0.40	0.5	0.23	0.34	0.33	0.41	0.34	0.41	0.30	0.41
		(0.26)		(0.18)		(0.22)		(0.13)		(0.13)		(0.16)
31	1.19	1.29	0.68	0.71	1.19	1.77	0.40	0.39	0.4	0.39	0.36	0.38
		(0.78)		(0.14)		(0.74)		(0.11)		(0.11)		(0.11)
32	0.44	0.44	0.40	0.38	0.48	0.51	0.17	0.31	0.17	0.32	0.29	0.36
		(0.31)		(0.33)		(0.4)		(0.14)		(0.13)		(0.20)
33	0.35	0.33	0.35	0.32	0.34	0.30	0.36	0.32	0.36	0.32	0.32	0.32
		(0.09)		(0.1)		(0.17)		(0.12)		(0.12)		(0.13)

Note: The table presents estimated elasticities of materials on output (measured in terms of quantity or revenue) from the estimation of Cobb Douglas (CD), or Translog (TL) production functions. The first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage ("None"). Translog specifications have heterogeneous elasticities within industries, with standard deviations presented in brackets. Industry codes refer to two-digit NACE codes. Industry names are provided in Table 2.

In line with the notion that firms face downward-sloping demand curves, we find that the elasticity of revenue with respect to materials is usually lower than the elasticity of quantity. For our preferred specification in the second column, where the elasticity comes from a translog production function using the full first stage and quantity data, we find higher average elasticities than the revenue-based counterpart in 16 out of 19 industries. On average, the quantity-based output elasticity exceeds the revenue-based elasticity by 34%. As derived by Bond et al. (2020), this implies that the true average markup of French manufacturing firms is 1.34.[37]

The estimated elasticities only modestly depend on the first-stage equation used. Notably, both for Cobb-Douglas and translog, we only find minor differences between the (average) estimated elasticities between the specification with the full first stage and the specification with no first stage at all. This is in line with the finding in Section 5 that the variance of the estimated η_{it} is low at around 3% of output. This means that the first-stage purging of the production function estimation procedure has only modest effect on the output used for the production function estimation. The columns with the "basic" first stage that does not have price or market share controls seem to have a slight downward bias in the estimated coefficients, but the difference in sector-level Cobb-Douglas elasticities and average translog elasticities is small.

The results in Table 8 show that industries differ significantly in the elasticity of output with respect to materials. Industries with low elasticities include NACE industry 13 (manufacturing of textile) and 15 (manufacturing of leather) while industries with high elasticities include NACE industry 31 (manufacturing of furniture) and 20 (manufacturing of chemicals). While Cobb-Douglas elasticities are usually close to the average elasticities under a translog production function, the table does show that there is sizable heterogeneity in elasticities across firms within industries. Standard deviations, which arise from non-zero values for the quadratic or interaction terms in the production function, are at least half the size of the average elasticity in the majority of specifications. Because we estimate markups by multiplying these elasticities with a firm's ratio of sales over materials, the dispersion of elasticities within sectors will cause a significant difference in the distribution of markups under a translog and a Cobb-Douglas production, as we show in the next section.

6.2 Markups

We next compute markups along the Hall (1986, 1988) equation using the estimated sector and firm-level elasticities. In the remaining analysis we focus on the log of markups. To treat for outliers we trim the bottom and top of the log markup distribution at the 1.5% level for each of

[37] This comparison uses the revenue production function with the full first stage, which includes controls for price. If the econometrician did not have price data, they would estimate the elasticity in the *Basic* first-stage column, which yields lower elasticities than for quantity in 15 out of 19 industries and implies a true average markup of 1.31. Note that the ratio of the average quantity and average revenue-based elasticities only needs to reflect the true markup in the absence of aggregate demand shocks.

Table 9: Overview - Log Markup Estimates

	Mean	St. Dev.	Median	10th Pct.	90th Pct.	Observations
			Cobb Douglas Production Function			
Quantity data						
Full first stage	0.29	0.52	0.24	-0.34	1.02	144686
Basic first stage	0.22	0.48	0.21	-0.36	0.85	144686
No first stage	0.28	0.49	0.22	-0.31	0.99	144686
Revenue data						
Full first stage	0.18	0.60	0.09	-0.48	0.95	144686
Basic first stage	0.04	0.60	0.07	-0.64	0.75	144686
No first stage	0.01	0.50	-0.02	-0.47	0.60	144686
			Translog Production Function			
Quantity data						
Full first stage	0.39	0.42	0.34	-0.04	0.91	144686
Basic first stage	0.31	0.28	0.29	-0.02	0.68	144686
No first stage	0.40	0.48	0.34	-0.07	0.97	144686
Revenue data						
Full first stage	0.14	0.20	0.13	-0.08	0.38	144686
Basic first stage	0.15	0.19	0.13	-0.07	0.38	144686
No first stage	0.14	0.18	0.12	-0.06	0.37	144686

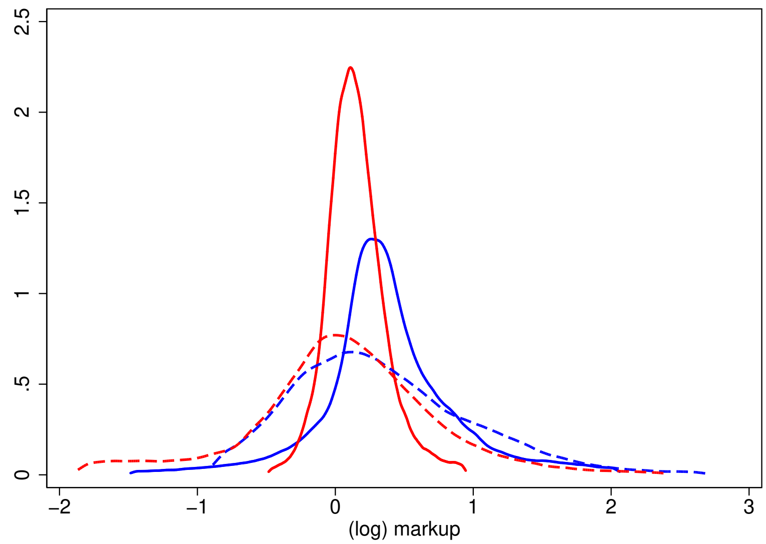
Note: All markups are expressed in log. Data for 2009-2018 from EAP-FARE.

the specifications. We focus on the non-trimmed sample to facilitate comparison. This leaves 144,686 firm-year observations. Summary statistics are provided in Table 9.

A clear pattern emerges from the table. First, the table shows that markups estimated from revenue data are consistently lower than markups estimated from quantity data. Average revenue-based markups are between 0.01 and 0.15 in logs and close to one in levels, in line with the bias described by Bond et al. (2020). The average of the log markup of the preferred specification, translog in quantity with the full first stage, is 0.39. This is in line with the markup implied by the ratio of the average revenue and quantity production function elasticities of 1.34, as discussed in the previous section. The average and median markups for the same specifications are similar for Cobb-Douglas and translog production functions, with the translog estimates being slightly higher for most specifications.

The second pattern is that the standard deviation of the estimated markups is significantly higher for the Cobb-Douglas production function. Standard deviations across markups from Cobb-Douglas production function specifications range from 0.48 to 0.60, while standard deviations for translog estimates range from 0.18 to 0.48. An illustration of this dispersion is provided in Figure 2, which plots the Kernel densities for a subset of the markup series. Red lines plot distributions of markup series that do not require pricing data and therefore use revenue for output and the basic first stage (without price controls), while blue lines plot distributions for the full first stage and quantity as the output variable. Solid lines plot distributions for the

Figure 2: Distribution of (Log) Markup by Production Function and Output Measure



Notes: The figure plots the distribution of markups (in logs) for four specifications. Solid lines denote markups from translog production functions, dashed lines from Cobb-Douglas production functions. Red lines denote markups from revenue data, blue lines from quantity data. The preferred series (translog production function, quantity data) is blue-solid.

translog production function, dashed lines for the Cobb-Douglas production function. The figure confirms that Cobb-Douglas markup estimates are significantly more dispersed.

To see what drives the difference in dispersion, note that Cobb-Douglas markups are given by:

$$\ln \widehat{\mu}_{it}^{CD} = \ln \widehat{\theta}^{M} + \ln \frac{P_{it}Y_{it}}{P_{it}^{M}M_{it}},$$

where $\widehat{\theta}^{M}$ is the estimated material elasticity at the sector level, such that the variance is:

$$Var(\ln \widehat{\mu}_{it}^{CD}) = Var\left(\ln \frac{P_{it}Y_{it}}{P_{it}^{M}M_{it}}\right).$$

Conversely, because the translog production function admits firm-specific elasticities, we have:

$$Var(\ln \widehat{\mu}_{it}^{TL}) = Var(\ln \widehat{\theta}_{it}^{M}) \cdot Var\left(\ln \frac{P_{it}Y_{it}}{P_{it}^{M}M_{it}}\right) + 2 \cdot Cov\left(\ln \frac{P_{it}Y_{it}}{P_{it}^{M}M_{it}}, \ln \widehat{\theta}_{it}^{M}\right),$$

where the first two right-hand terms are positive. Hence, the lower dispersion of markups in the translog specification implies that there is a negative correlation between a firm's revenue-over-materials share and the elasticity of its output with respect to materials.

The overstatement of markup dispersion in the Cobb-Douglas specifications is particularly important because researchers frequently assume a Cobb-Douglas production function to 'prevent' having to estimate a production function: when studying log markups, by taking indus-

Table 10: Costs of Markup Dispersion by Production Function Estimate

	St. Dev. of Log Markups	Dispersion costs
Translog - Quantity	0.42	8.15%
Translog - Revenue	0.20	1.78%
Cobb-Douglas - Quantity	0.52	11.8%
Cobb-Douglas - Revenue	0.60	18.5%

Notes: Example calculation of how the costs of markup dispersion change when using alternative production function estimates using the formula in Peters (2020). Data for the EAP-FARE sample (2009-2018). Dispersion costs are expressed as $(1 - \mathcal{M}) \cdot 100\%$. Markups from quantity data are estimated using the full first stage that includes price as a control, markups from revenue data are estimated using the basic first stage that does not include price as a control.

try fixed effects to correct for θ^M, researchers can analyze within-industry markup dispersion simply from observing revenue and material expenditures. Figure 2, however, shows that the Cobb-Douglas assumption is rejected in the French data. This is important for a number of applications, for example when analyzing the effect of heterogeneous markups on allocative efficiency and productivity. The idea is that firms with higher markups produce inefficiently little because they raise prices above marginal costs. As the Cobb-Douglas production function estimates overstate the degree of markup dispersion, a researcher relying on these estimates would therefore overstate the degree of misallocation in the economy.[38]

By how much a researcher would overstate the costs of markups when using Cobb-Douglas estimates would depend on assumptions about the demand system and on the drivers of markup dispersion. By means of a back-of-the-envelope exercise, we perform the simple misallocation cost calculating in Peters (2020) for the case of innovation-driven markup dispersion with a Cobb-Douglas aggregator. He shows that the ratio of true aggregate productivity and productivity under allocative efficiency is given by:

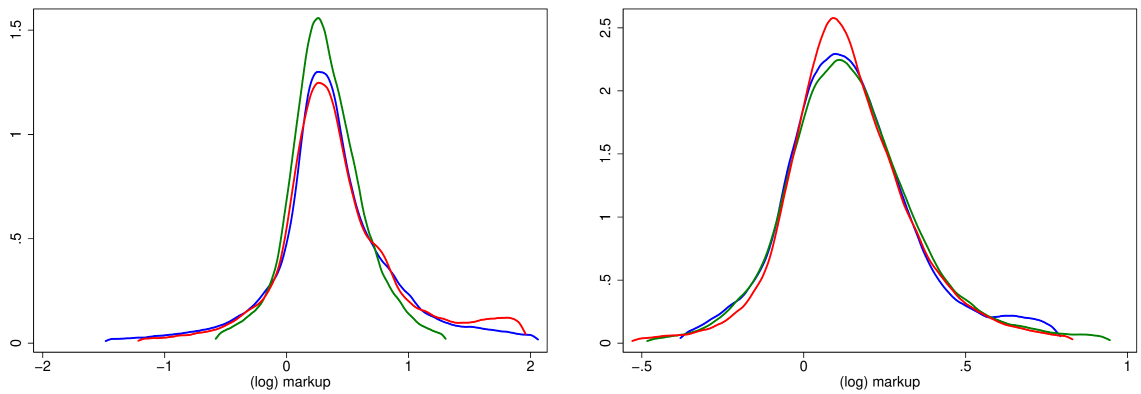
$$\mathcal{M} = \frac{\exp\left(\sum_{i \in I} \ln \mu_i^{-1}\right)}{\sum_{i \in I} \mu_i^{-1}},$$

where I denotes the set of all firms. The denominator and numerator are equal if firms have homogeneous markups, such that aggregate productivity equals its efficient benchmark. Table 10 presents the results. Assuming that the translog quantity markups are correct, the true reduction in aggregate total factor productivity because of markup dispersion is 8.15%. A researcher that would use Cobb-Douglas estimates from revenue data would measure the cost of dispersion at 18.5%, an overstatement of 126%.

A third pattern in Table 9 is that the distributions of markups across first-stage specifications are similar when holding the production function and output variable constant. For translog, nearly all published moments are within five log-points of eachother. An illustration of how

[38] A researcher could equally *underestimate* the costs of markup dispersion in case the ratio of sales over material spending is positively correlated with the output elasticity of materials. The point here is that the Cobb-Douglas production function assumption is not without loss of generality.

Figure 3: Distribution of (Log) Markup by Output Variable and First Stage Specification



(a) Quantity

(b) Revenue

Notes: The figure plots the distribution of markups (in logs). Blue lines present markups estimated with a first stage that includes price and market share controls. Green lines present specifications that exclude price and market share controls. Red lines present results from specifications without a first stage.

similar these distributions are is provided in Figure 3. For the translog production function it plots the full (blue), basic (green) and no (red) first stage specifications. The left figure plots the kernel densities for quantity while the right figure plots them for revenue. While there are some differences, especially at the tails of the distribution, the distributions are overall very similar in terms of mean, median and standard deviation. It therefore seems hat the choice of the functional form of the production function and the choice of the output variable has a much larger effect on the estimated markups than the exact specification of the first stage.

6.3 Markups: Correlations

We next assess how markup estimates correlate. First we measure correlations between markups from alternative specifications and then analyze whether the relationship between markups and key variables depend on the measure of the markup that is deployed.

The correlation between markups from various specifications is presented in Table 11. The top panel in the table presents correlations for markups from Cobb-Douglas production functions while the bottom panel in the table presents correlations from the translog production function. For the Cobb-Douglas production function, the table generally shows high correlations across specifications. With few exceptions correlations exceed 0.5. Correlations are particularly high for markups based on the same output variable. For quantity, for example, the correlation between markups estimated using the full first stage and no first stage is 0.98. Correlations across output variables are lower. The correlation between markups using the full first stage from quantity and revenue data, for example, is 0.6. While that is lower than correlations within the same output variable, it still implies that revenue-based markups are informative

Table 11: Correlations across Specifications - Log Markups

	Full - Q	Full - R	Basic - Q	Basic - R	None - Q	None - R
	Cobb-Douglas Production Function					
Full first stage - Quantity data	1.00	0.60	0.80	0.52	0.98	0.66
Full first stage - Revenue data	0.60	1.00	0.64	0.19	0.60	0.55
Basic first stage - Quantity data	0.80	0.64	1.00	0.54	0.87	0.77
Basic first stage - Revenue data	0.52	0.19	0.54	1.00	0.53	0.48
No first stage - Quantity data	0.98	0.60	0.87	0.53	1.00	0.68
No first stage - Revenue data	0.66	0.55	0.77	0.48	0.68	1.00
	Translog Production Function					
Full first stage - Quantity data	1.00	0.30	0.77	0.34	0.79	0.36
Full first stage - Revenue data	0.30	1.00	0.47	0.71	0.23	0.56
Basic first stage - Quantity data	0.77	0.47	1.00	0.45	0.71	0.45
Basic first stage - Revenue data	0.34	0.71	0.45	1.00	0.23	0.59
No first stage - Quantity data	0.79	0.23	0.71	0.23	1.00	0.25
No first stage - Revenue data	0.36	0.56	0.45	0.59	0.25	1.00

Note: Each cell presents the pairwise correlation between the markup in the row and the column header. The first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage. All markups are expressed in log. Data for 2009-2018 from EAP-FARE.

about a firm's actual markup.

Correlations across markups from translog production function estimates are generally lower. This is expected, as variation in markups for the same firm across specifications under Cobb Douglas only comes from sector-level production function estimates. For the translog production function each firm's estimated output elasticity changes, causing greater differences across specifications. Nevertheless we find consistently positive correlations across the specifications. Like before, these correlations are particularly strong when comparing markups with production functions that use the same output variable. For quantity, for example, the correlation between markups estimated using the full first stage and no first stage is now 0.79, while the correlation between markups from full-first stage quantity and revenue data is now 0.3.

Note that the correlation between markups from various specifications is significantly greater when analyzing log-differences (Table 12). For Cobb-Douglas the correlation across all specifications equals unity trivially, because the output elasticity of materials is constant within the industry and constant over time. There is also a marked increase, however, in the correlation between translog markups. In particular, there is a stronger correlation in first differences between translog markups estimated with revenue and quantity, with the full-first stage specification now having a 0.61 pairwise correlation coefficient. Researchers interested in changes rather than levels of markups may face lower bias from the use of revenue data in production function estimations than researchers interested in the level of the markup.

In a final exercise, we assess whether relationships between markups and key variables depends on the markup specification. To do so, we estimate a series of simple regressions of the

Table 12: Correlations across Specifications for Log-Differenced Markups

	Full - Q	Full - R	Basic - Q	Basic - R	None - Q	None - R
	Cobb-Douglas Production Function					
Full first stage - Quantity data	1.00	1.00	1.00	1.00	1.00	1.00
Full first stage - Revenue data	1.00	1.00	1.00	1.00	1.00	1.00
Basic first stage - Quantity data	1.00	1.00	1.00	1.00	1.00	1.00
Basic first stage - Revenue data	1.00	1.00	1.00	1.00	1.00	1.00
No first stage - Quantity data	1.00	1.00	1.00	1.00	1.00	1.00
No first stage - Revenue data	1.00	1.00	1.00	1.00	1.00	1.00
	Translog Production Function					
Full first stage - Quantity data	1.00	0.61	0.75	0.61	0.60	0.61
Full first stage - Revenue data	0.61	1.00	0.82	0.96	0.62	0.90
Basic first stage - Quantity data	0.75	0.82	1.00	0.81	0.70	0.79
Basic first stage - Revenue data	0.61	0.96	0.81	1.00	0.59	0.90
No first stage - Quantity data	0.60	0.62	0.70	0.59	1.00	0.60
No first stage - Revenue data	0.61	0.90	0.79	0.90	0.60	1.00

Note: Each cell presents the pairwise correlation between the markup in the row and the column header. The first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage. All markups are expressed in log difference. Data for 2009-2018 from EAP-FARE.

following kind for each markup specification *s*:

$$x_{it} = \chi \ln \widehat{\mu}_{it}^{s} + \varphi_i + \psi_t + \epsilon_{it},$$

where respectively φ_i and ψ_t denote firm and time fixed effects, and where x_{it} denotes some variable of interest. We estimate this regression using a firm's profit rate (defined as the ratio of operating profits over sales), labor share (defined as the ratio of its wagebill over sales), its material cost share (defined as the ratio of materials purchased over sales), and its market share (defined as its share in revenue at the 5-digit sector level) as dependent variables. Our aim is not to causally estimate the relationship between these variables and markups, but rather to see how the conditional correlation between these variables and markups depends on how the production function was estimated. We only include translog markup estimates, because the relationship between the explanatory variable is identical for all Cobb-Douglas estimates given the inclusion of firm fixed effects.

Results are presented in Table 13. Each row presents regression coefficients for a particular explanatory variable (described in italics), while each column contains results for a specific markup specification. Before describing differences across specifications, note that all relationships in the table run in the expected direction. Firms with higher markup estimates are more profitable, have lower labor shares, lower material shares, and greater market shares. This is the case irrespective of whether revenue or quantity data was used to estimate the production function elasticities, and the relationships are all significant at the 1% level. Looking more carefully at the specifications, we see that estimated βs do differ across specifications, both when

Table 13: Relation between Markup and Explanatory Variables by Markup Specification

	Quantity			Revenue		
	Full	Basic	None	Full	Basic	None
Profit Rate						
Estimated χ	0.190***	0.424***	0.237***	0.666***	0.632***	0.580***
	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)	(0.006)
R-squared	0.128	0.300	0.175	0.491	0.464	0.437
Labor Share						
Estimated χ	-0.024***	-0.111***	-0.0923***	-0.190***	-0.158***	-0.183***
	(0.003)	(0.004)	(0.003)	(0.004)	(0.004)	(0.004)
R-squared	0.015	0.061	0.076	0.110	0.083	0.118
Materials Share						
Estimated χ	-0.142***	-0.292***	-0.128***	-0.403***	-0.367***	-0.304***
	(0.003)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
R-squared	0.192	0.376	0.143	0.472	0.415	0.321
Market Share (%)						
Estimated χ	0.0792***	0.124***	0.101***	0.184***	0.160***	0.162***
	(0.009)	(0.019)	(0.013)	(0.021)	(0.020)	(0.017)
R-squared	0.006	0.007	0.007	0.008	0.007	0.007
Firm Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Time Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	144,686	144,686	144,686	144,686	144,686	144,686

Note: Each entry gives the OLS regression coefficient with the cursive variable as the explanatory variable and the markup series in the column header as the explanatory variable. Markups are estimated using two stage regression where the first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage ("None"). All markups are expressed in logs. Each markup series is trimmed at 1.5% tails by specification and the sample consists of the subset of firm-years for which all specifications are available after trimming. Explanatory variables are winsorized at 1.5% tails. Firm-clustered standard errors in parentheses. *** denotes significance at the 1% level. Data for 2009-2018 from EAP-FARE.

changes are made to the first stage or when quantity or revenue is used. The estimated βs tend to be smaller for quantity-based markups than for revenue-based markups. This is in line with the finding in Figure 2 that there is more dispersion in the quantity-based markup estimates; a higher variance of the markup mechanically reduces the estimated βs holding everything else equal. Overall, however, the results in Table 13 suggest that relationships between markups and key relationships are qualitatively robust to using imperfect first stage regressions or revenue-based markup estimates. This further supports our derivation in Section 2 that these estimates do contain useful information about a firm's true markup.

7 Conclusion

This paper provides an assessment of the validity of the ratio estimator of firm-level markups. We start by deriving the conditions under which the commonly used two-stage iterative GMM estimator is able to consistently estimate the parameters of the production function, in a simplified framework and relaxing the assumption that firms are price setters. We then use the simple framework to look at three common critiques and practical difficulties of using the ratio estimator (and the iterative GMM production function estimates) to gauge markups: the assumption of a Cobb-Douglas production function, the use of revenue instead of quantity to measure firm output, and the improper specification of the first-stage regression. We confirm the resulting insights with simulations from a rich macro model and empirical data on prices and production for French manufacturing firms.

We find that, while the average markup is estimated reasonably well under Cobb-Douglas, the dispersion of markups is not. In the French data, a researcher that measures misallocation costs of markup dispersion would overstate costs by 126% if relying on a Cobb-Douglas production function. Conversely, we find that the use of revenue rather than quantity to estimate production functions affects the level of the estimated markups, but has modest effects on dispersion. The correlation between markups from quantity and revenue data ranges from 0.3 to 0.7 in log-levels and is at least 0.59 in log-differences. The correlation between various markup estimates and variables such as market share, profitability and the labor share is also similar across the use of revenue or quantity data. We find that empirical estimates of the markups do not depend strongly on the specification of the first-stage regression in the two-stage procedure.

Practically, we conclude that if a researcher is faced with imperfect data it depends on individual applications whether the analysis can proceed. Optimally, production functions for markups should be estimated with quantity rather than revenue data. In the absence of data on prices however, researchers that are interested in the dispersion or correlations of markups should hesitate to assume a Cobb-Douglas production function. Conversely, in applications where researchers are interested in the average level of the markup, revenue data may not be appropriate. Revenue data may be used to estimate trends of markups (as differences over time are a part of dispersion), provided the researcher is willing to assume that the production function parameters do not change over time.

References

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer, "Identification Properties of Recent Production Function Estimators," *Econometrica*, 2015, *83* (6), 2411–2451.

Ackerberg, Daniel, C. Lanier Benkard, Steven Berry, and Ariel Pakes, "Econometric Tools for Analyzing Market Outcomes," in J.J. Heckman and E.E. Leamer, eds., *Handbook of Econometrics*, Vol. 6 of *Handbook of Econometrics*, Elsevier, January 2007, chapter 63.

Aghion, Philippe, Antonin Bergeaud, Timo Boppart, Peter J. Klenow, and Huiyu Li, "A Theory of Falling Growth and Rising Rents," Working Paper Series 2019-11, Federal Reserve Bank of San Francisco March 2019.

_ , Nick Bloom, Richard Blundell, Rachel Griffith, and Peter Howitt, "Competition and innovation: An inverted-U relationship," *The quarterly journal of economics*, 2005, *120* (2), 701–728.

Atkeson, Andrew and Ariel Burstein, "Pricing-to-Market, Trade Costs, and International Relative Prices," *American Economic Review*, December 2008, *98* (5), 1998–2031.

Autor, David, David Dorn, Lawrence F Katz, Christina Patterson, and John Van Reenen, "The fall of the labor share and the rise of superstar firms," *The Quarterly Journal of Economics*, 2020, *135* (2), 645–709.

Baqaee, David Rezza and Emmanuel Farhi, "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem," *Econometrica*, July 2019, *87* (4), 1155–1203.

Barkai, Simcha, "Declining labor and capital shares," Technical Report, London Business School 2017.

Basu, Susanto, "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence," *Journal of Economic Perspectives*, August 2019, *33* (3), 3–22.

_ and John G Fernald, "Returns to Scale in U.S. Production: Estimates and Implications," *Journal of Political Economy*, April 1997, *105* (2), 249–83.

Blundell, Richard and Stephen Bond, "GMM estimation with persistent panel data: an application to production functions," *Econometric reviews*, 2000, *19* (3), 321–340.

Bond, Steve, Arshia Hashemi, Greg Kaplan, and Piotr Zoch, "Some Unpleasant Markup Arithmetic: Production Function Elasticities and their Estimation from Production Data," Working Paper 27002, National Bureau of Economic Research April 2020.

Bornstein, Gideon, "Entry and profits in an aging economy: The role of consumer inertia," Technical Report, Working Paper 2018.

Brand, James, "Estimating Productivity and Markups Under Imperfect Competition."," 2019.

Burstein, Ariel, Vasco M. Carvalho, and Basile Grassi, "Bottom-up Markup Fluctuations," NBER Working Papers 27958, National Bureau of Economic Research, Inc October 2020.

Calligaris, Sara, Chiara Criscuolo, and Luca Marcolin, "Mark-ups in the digital era," 2018.

Cavenaile, Laurent, Murat Alp Celik, and Xu Tian, "Are markups too high? competition, strategic innovation, and industry dynamics," *Competition, Strategic Innovation, and Industry Dynamics (August 1, 2019)*, 2019.

Collard-Wexler, Allan and Jan De Loecker, "Production Function Estimation and Capital Measurement Error," Technical Report, National Bureau of Economic Research 2020.

Crouzet, Nicolas and Janice C Eberly, "Understanding weak capital investment: The role of market concentration and intangibles," Technical Report, National Bureau of Economic Research 2019.

De Loecker, Jan and Frederic Warzynski, "Markups and Firm-Level Export Status," *American Economic Review*, May 2012, *102* (6), 2437–71.

Decker, Ryan A, John Haltiwanger, Ron S Jarmin, and Javier Miranda, "Changing business dynamism and productivity: Shocks versus responsiveness," *American Economic Review*, 2020, *110* (12), 3952–90.

Doraszelski, Ulrich and Jordi Jaumandreu, "Using Cost Minimization to Estimate Markups," 2019.

__ **and** __ , "The inconsistency of De Loecker and Warzynski's (2012) method to estimate markups and some robust alternatives," 2020.

Díez, Federico, Jiayue Fan, and Carolina Villegas-Sanchez, "Global Declining Competition," CEPR Discussion Papers 13696, C.E.P.R. Discussion Papers April 2019.

Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu, "How Costly Are Markups?," Working Paper 24800, National Bureau of Economic Research July 2018.

Eggertsson, Gauti B, Jacob A Robbins, and Ella Getz Wold, "Kaldor and Piketty's Facts: The Rise of Monopoly Power in the United States," Working Paper 24287, National Bureau of Economic Research February 2018.

Gandhi, Amit, Salvador Navarro, and David A Rivers, "On the identification of gross output production functions," *Journal of Political Economy*, 2020, *128* (8), 2973–3016.

Grassi, Basile, "IO in I-O: Size, Industrial Organization, and the Input-Output NetworkMake a Firm Structurally Important," Working Papers 619, IGIER (Innocenzo Gasparini Institute for Economic Research), Bocconi University 2018.

Grullon, Gustavo, Yelena Larkin, and Roni Michaely, "Are US industries becoming more concentrated?," *Review of Finance*, 2019, *23* (4), 697–743.

Gutiérrez, Germán and Sophie Piton, "Revisiting the global decline of the (non-housing) labor share," *American Economic Review: Insights*, 2020, *2* (3), 321–38.

Gutiérrez, Germán and Thomas Philippon, "Declining Competition and Investment in the U.S," NBER Working Papers 23583, National Bureau of Economic Research, Inc July 2017.

__ **and** __ , "How EU Markets Became More Competitive Than US Markets: A Study of Institutional Drift," Working Paper 24700, National Bureau of Economic Research June 2018.

Hall, Robert E, "Market structure and macroeconomic fluctuations," *Brookings papers on economic activity*, 1986, *1986* (2), 285–338.

Hall, Robert E., "The Relation between Price and Marginal Cost in U.S. Industry," *Journal of Political Economy*, 1988, *96* (5), 921–947.

Hershbein, Brad, Claudia Macaluso, and Chen Yeh, "Monopsony in the U.S. Labor Market," *Working Power*, 2021.

Hsieh, Chang-Tai and Peter J. Klenow, "Misallocation and Manufacturing TFP in China and India," *The Quarterly Journal of Economics*, November 2009, *124* (4), 1403–1448.

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

Kehrig, Matthias and Nicolas Vincent, "The micro-level anatomy of the labor share decline," *The Quarterly Journal of Economics*, 2021, *136* (2), 1031–1087.

Kimball, Miles S., "The Quantitative Analytics of the Basic Neomonetarist Model," *Journal of Money, Credit and Banking*, 1995, *27* (4), 1241–1277.

Klenow, Peter J. and Jonathan L. Willis, "Real Rigidities and Nominal Price Changes," *Economica*, July 2016, *83* (331), 443–472.

Klette, Tor Jakob and Zvi Griliches, "The inconsistency of common scale estimators when output prices are unobserved and endogenous," *Journal of applied econometrics*, 1996, *11* (4), 343–361.

Levinsohn, James and Amil Petrin, "Estimating Production Functions Using Inputs to Control for Unobservables," *Review of Economic Studies*, 2003, *70* (2), 317–341.

Loecker, Jan De, Jan Eeckhout, and Gabriel Unger, "The Rise of Market Power and the Macroeconomic Implications*," *The Quarterly Journal of Economics*, 01 2020, *135* (2), 561–644.

__ , Pinelopi K. Goldberg, Amit K. Khandelwal, and Nina Pavcnik, "Prices, Markups, and Trade Reform," *Econometrica*, 2016, *84* (2), 445–510.

Meier, Matthias and Timo Reinelt, *Monetary Policy, Markup Dispersion, and Aggregate TFP*, Working Paper, 2020.

Morlacco, Monica, "Market Power in Input Markets: Theory and Evidence from French Manufacturing," Technical Report 2019.

Neiman, Brent and Joseph S Vavra, "The rise of niche consumption," Technical Report, National Bureau of Economic Research 2019.

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

Peters, Michael, "Heterogeneous markups, growth, and endogenous misallocation," *Econometrica*, 2020, *88* (5), 2037–2073.

Ridder, M. De, "Market Power and Innovation in the Intangible Economy," Cambridge Working Papers in Economics 1931, Faculty of Economics, University of Cambridge March 2019.

Syverson, Chad, "Macroeconomics and Market Power: Context, Implications, and Open Questions," *Journal of Economic Perspectives*, August 2019, *33* (3), 23–43.

Traina, James, "Is Aggregate Market Power Increasing? Production Trends Using Financial Statements," Working Paper 17, Stigler Cente February 2018.

A Analytical Framework Derivations

A.1 Finite Sample Estimator and its Asymptotic Variance

In this section, we derive the estimator for a finite sample. We also use this derivation to compute the asymptotic variance of the GMM estimator. First, let us defined the estimator for a finite sample.

Definition: GMM estimator is $\widehat{\alpha}$ such that $\sum_{i,t} \widehat{\omega}_{it} v_{it-1} = 0$ with $\widehat{\omega}_{it} = y_{it} - \widehat{\alpha} v_{it} = (\alpha - \widehat{\alpha}) v_{it} + \omega_{it}$.

Second, to solve for the estimator, we need to find the value of $\widehat{\alpha}$ such that $\sum_{i,t} \widehat{\omega}_{it} v_{it-1} = (\alpha - \widehat{\alpha}) \sum_{i,t} v_{it} v_{it-1} + \sum_{i,t} \omega_{it} v_{it-1} = 0$. As long as $\sum_{i,t} v_{it} v_{it-1} \neq 0$, the unique $\widehat{\alpha}$ that solve this equation is

$$\widehat{\alpha} = \alpha + \frac{\sum_{i,t} \omega_{it} v_{it-1}}{\sum_{i,t} v_{it} v_{it-1}}$$

whose limit is α when the sample size increase where since $\mathbb{E}[\omega_{it} v_{it-1}] = 0$.

Finally, let us derive the asymptotic variance of the GMM estimator. Using the (finite sample) expression of the estimator, we have

$$\sqrt{n}(\widehat{\alpha} - \alpha) = \frac{\sqrt{n}\frac{1}{n} \sum_{i,t} \omega_{it} v_{it-1}}{\frac{1}{n} \sum_{i,t} v_{it} v_{it-1}}.$$

By the (weak) law of large number, $\frac{1}{n} \sum_{i,t} v_{it} v_{it-1} \xrightarrow{p} \mathbb{E}[v_{it} v_{it-1}]$, and, by the central limit theorem, $\sqrt{n}\frac{1}{n} \sum_{i,t} \omega_{it} v_{it-1} \xrightarrow{d} \mathcal{N}\left(0, \mathbb{E}\left[\omega_{it}^2 v_{it-1}^2\right]\right)$. The Slutsky theorem implies $\sqrt{n}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}\left(0, \frac{\mathbb{E}\left[\omega_{it}^2 v_{it-1}^2\right]}{\mathbb{E}[v_{it} v_{it-1}]^2}\right)$, that is,

$$\mathbb{V}ar\left[\widehat{\alpha}\right] \sim \frac{\mathbb{E}\left[\omega_{it}^2\right] \mathbb{E}\left[v_{it-1}^2\right]}{\sqrt{n}\mathbb{E}[v_{it} v_{it-1}]^2}.$$

A.2 Global results on GMM estimator with AR(1) in productivity

The GMM estimator with AR(1) productivity (Definition 4) is characterized by the system of equations:

$$\begin{cases} \mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] = 0 \\ \mathbb{E}\left[\widehat{\xi}_{it} \widehat{\omega}_{it-1}\right] = 0 \end{cases} \iff \begin{cases} \mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] = 0 \\ (\alpha - \widehat{\alpha})\mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] + \mathbb{E}\left[\widehat{\xi}_{it} \omega_{it-1}\right] = 0 \end{cases} \iff \begin{cases} \mathbb{E}\left[\widehat{\xi}_{it} v_{it-1}\right] = 0 \\ \mathbb{E}\left[\widehat{\xi}_{it} \omega_{it-1}\right] = 0 \end{cases}$$

$$\iff \begin{cases} \mathbb{E}\left[\xi_{it} v_{it-1}\right] + (\alpha - \widehat{\alpha})\mathbb{E}\left[(v_{it} - \rho v_{it-1}) v_{it-1}\right] + (\rho - \widehat{\rho})\mathbb{E}\left[\omega_{it-1} v_{it-1}\right] + (\alpha - \widehat{\alpha})(\rho - \widehat{\rho})\mathbb{E}\left[v_{it-1}^2\right] = 0 \\ \mathbb{E}\left[\xi_{it} \omega_{it-1}\right] + (\alpha - \widehat{\alpha})\mathbb{E}\left[(v_{it} - \rho v_{it-1}) \omega_{it-1}\right] + (\rho - \widehat{\rho})\mathbb{E}\left[\omega_{it-1}^2\right] + (\alpha - \widehat{\alpha})(\rho - \widehat{\rho})\mathbb{E}\left[v_{it-1} \omega_{it-1}\right] = 0 \end{cases}$$

$$\iff \begin{cases} g + aX + bY + cXY = 0 \\ h + dX + eY + fXY = 0 \end{cases}$$

where $X = \alpha - \widehat{\alpha}$, $Y = \rho - \widehat{\rho}$, and, $a = \mathbb{E}\left[(v_{it} - \rho v_{it-1}) v_{it-1}\right]$, $b = \mathbb{E}\left[\omega_{it-1} v_{it-1}\right]$, $c = \mathbb{E}\left[v_{it-1}^2\right]$, $d = \mathbb{E}\left[(v_{it} - \rho v_{it-1}) \omega_{it-1}\right]$, $e = \mathbb{E}\left[\omega_{it-1}^2\right]$, $f = \mathbb{E}\left[v_{it-1} \omega_{it-1}\right] = b$, $g = \mathbb{E}\left[\xi_{it} v_{it-1}\right]$, $h = \mathbb{E}\left[\xi_{it} \omega_{it-1}\right]$.

Let us look at the asymptotic where $g = 0$ and $h = 0$. Assuming $c \neq 0$, we get

$$\left\{\begin{array}{l} aX + bY + cXY = 0 \\ dX + eY + fXY = 0 \end{array}\right. \iff \left\{\begin{array}{l} X = 0 \\ Y = 0 \end{array}\right. \text{ or } \left\{\begin{array}{l} X = -\frac{bd-ae}{cd-af} \\ Y = \frac{bd-ae}{ce-bf} \end{array}\right. \text{ if } cd - af \neq 0 \text{ and } ce - bf \neq 0.$$

It follows that there is two global solutions for the GMM estimator with AR(1):

$$\left\{\begin{array}{l} \widehat{\alpha} = \alpha \\ \\ \widehat{\rho} = \rho \end{array}\right. \text{ or } \left\{\begin{array}{l} \widehat{\alpha} = \alpha - \frac{bd-ae}{cd-af} = \alpha - \sqrt{\frac{\mathbb{V}ar[\omega_{it-1}]}{\mathbb{V}ar[v_{it-1}]}} \frac{Corr(\widetilde{v}_{it}, v_{it-1}) - Corr(\widetilde{v}_{it}, \omega_{it-1}) Corr(\omega_{it-1}, v_{it-1})}{Corr(\widetilde{v}_{it}, \omega_{it-1}) - Corr(\widetilde{v}_{it}, v_{it-1}) Corr(\omega_{it-1}, v_{it-1})} \\ \\ \widehat{\rho} = \rho + \frac{bd-ae}{ce-bf} = \rho + \sqrt{\frac{\mathbb{V}ar[\widetilde{v}_{it}]}{\mathbb{V}ar[v_{it-1}]}} \frac{Corr(\widetilde{v}_{it}, v_{it-1}) - Corr(\widetilde{v}_{it}, \omega_{it-1}) Corr(\omega_{it-1}, v_{it-1})}{1 - Corr(\omega_{it-1}, v_{it-1})^2} \end{array}\right.$$

where $\widetilde{v}_{it} \equiv v_{it} - \rho v_{it-1} = \frac{1}{1-\alpha} \left(\xi_{it} + mc_{it} - \rho mc_{it-1} + w_t - \rho w_{t-1}\right)$.[39] The GMM estimator admits (exactly) two possible solutions. One solution giving the true value of the parameters, while the second solution giving a bias estimate of the true parameters. However, if $\mathbb{V}ar[v_{it-1}]$ is large compared to $\mathbb{V}ar[\omega_{it-1}]$ and $\mathbb{V}ar[\widetilde{v}_{it}]$, that is, their ratio goes to infinity while keeping fix the correlation structure, then there a unique solution for $\widehat{\alpha}$ and $\widehat{\rho}$. To conclude, if there is enough variation in the data, the GMM estimator is identified.

A.3 Full Proof

TBD

B Potential Bias

B.1 Approximation of Demand System

In this appendix, we show how to approximate the demand system specify by $Y = D(P)$ or $P = D^{-1}(Y)$. Note that both of these demand system allows for differentiated goods across firms as described in **?**. For the former case, let us defined the function $D_{it}(P)$ such that $Y_{it} = D_{it}(P)$. Around some symmetric equilibrium, (P_0^*, Y_0^*), at the first-order we have, for all i, t

$$y_{it} = \log Y_{it} - \log Y_0^* \approx \sum_{jt} \frac{\partial \log D_{it}}{\partial \log P_{jt}} \left(\log P_{jt} - \log P_0^*\right) = \sum_{jt} J_{ijt} p_{jt}$$

where the matrix whose element are J_{ijt} is the Jacobian of the log of the demand D. Inverting this system of equation yields that for all i, $p_{it} = \sum_{jt} d_{ijt} y_{jt}$ where d_{ijt} are the element of the inverse of the Jacobian matrix of the (log) demand D. For this case, when the demand is specify by $Y = D(P)$, we need to assume that the Jacobian of $\log D$ is invertible.

For the case where the demand is given by the inverse demand directly, $P = D^{-1}(Y)$, let us

[39] Note that $Corr(\widetilde{v}_{it}, \omega_{it-1}) = Corr(mc_{it} - \rho mc_{it-1} + w_t - \rho w_{t-1}, \omega_{it-1})$. Intuitively, if input price and marginal cost ($= P_{it}/\mu_{it}$) are uncorrelated with past value of productivity, this correlation will be equal to zero.

defined the function D_{it}^{-1} such that $P_{it} = D_{it}^{-1}(Y)$. A first-order approximation around a symmetric equilibrium (P_0^*, Y_0^*) yields

$$p_{it} = \log P_{it} - \log P_0^* \approx \sum_{jt} \frac{\partial \log D_{it}^{-1}}{\partial \log Y_{jt}} (\log Y_{jt} - \log Y_0^*) = \sum_{jt} d_{ijt} y_{jt},$$

where, here, the d_{ijt} are the element of the Jacobian matrix of the (log) inverse demand D^{-1}.

These formulation are useful when deriving the markup of firms of static oligopolistic Cournot or Bertrand competition. Under Bertrand, that is when firms takes other firm's prices as given, the profit of firm i at time t can be written as $\Pi_{it} = P_{it}Y_{it} - C_{it}(Y_{it}) = P_{it}D_{it}(P) - C_{it}(D_{it}(P))$ where $C_{it}(Y_{it})$ is the total cost of producing Y_{it} units. Under Bertrand, firms maximize their profit by setting their price P_{it} taking others' price as given. The first-order condition of this profit maximization problem yields that the markup is $\mu_{it} \equiv \frac{P_{it}}{\frac{\partial C_{it}}{\partial Y_{it}}} = \left(1 + \left(\frac{\partial \log D_{it}}{\partial \log P_{it}}\right)^{-1}\right)^{-1}$. Similarly, under Cournot, the profit of firm i at time t can be written as $\Pi_{it} = P_{it}Y_{it} - C_{it}(Y_{it}) = D_{it}^{-1}(Y)Y_{it} - C_{it}(Y_{it})$. Under Cournot, firms choose their quantity, taking other firm's quantity as given, which implies that the markup is $\mu_{it} \equiv \frac{P_{it}}{\frac{\partial C_{it}}{\partial Y_{it}}} = \left(1 + \frac{\partial \log D_{it}^{-1}}{\partial \log Y_{it}}\right)^{-1}$. To conclude, in most static oligopolistic competition model the firm-level markup can be written as $\mu_{it} = (1 + d_{iit})^{-1}$.[40]

B.2 Revenue Markup and Translog Production Function

We next compare markups from revenue and quantity production functions in a more general framework with a translog production function. The main intuition remains valid: the bias of the estimator on revenue data is equal to the *average* demand elasticity among firms sharing the same production function.

Assume that the production function is as in section 2.2.1, that is, $y_{it} = \alpha v_{it} + \beta v_{it}^2 + \omega_{it}$ while we maintain the other assumptions of our baseline framework. Let us study the bias implied by the use of revenue data in place of quantity data. Following the same logic as above, the coefficients of the production function estimated on revenue are such that

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \begin{pmatrix} \alpha \\ \beta \end{pmatrix} + V^{-1} \begin{pmatrix} \mathbb{E}[p_{it}v_{it-1}] \\ \mathbb{E}[p_{it}v_{it-1}^2] \end{pmatrix}, \qquad \text{with} \qquad V = \begin{pmatrix} \mathbb{E}[v_{it}v_{it-1}] & \mathbb{E}[v_{it}^2 v_{it-1}] \\ \mathbb{E}[v_{it}v_{it-1}^2] & \mathbb{E}[v_{it}^2 v_{it-1}^2] \end{pmatrix}.$$

As in the Cobb-Douglas case, these estimates are biased. The above equation is the translog equivalent of equation (2) where the correlation of the instruments (lagged variable inputs and lagged variable inputs squared) with the output price is the case of the bias.

In the case of translog production function, the true markup is such that $\mu_{it} = (\alpha +$

[40] Under Cournot, we always have $\mu_{it} = (1 + d_{iit})^{-1}$, while under Bertrand, we further need to assume that d_{iit} the diagonal term of the Jacobian matrix of the (log) inverse demand D^{-1} is equal to $\left(\frac{\partial \log D_{it}}{\partial \log P_{it}}\right)^{-1}$.

$2\beta \log V_{it})\frac{P_{it}Y_{it}}{W_t V_{it}}$, and, the revenue markup is thus $\widehat{\mu}_{it}^R = \frac{\widehat{\alpha}+2\widehat{\beta}\log V_{it}}{\alpha+2\beta\log V_{it}}\mu_{it}$. As pointed out by Bond et al. (2020) and as in the Cobb-Douglas case, if we assume homogeneous inverse demand elasticities among firms in the sample, that is for all i we have $p_{it} = -\gamma y_{it}$, the revenue markup is equal to one.[41] However, in general the revenue markup is different from one and contains information on the true markup. To see this formally, we assume again that inverse demand elasticities are heterogeneous among firms, such that for all i by $p_{it} = -d_{iit} y_{it}$ where there is at least one pair (i, j) such that $d_{iit} \neq d_{jjt}$. As above, the true markup is given by $\mu_{it} = (1 - d_{iit})^{-1}$. In this heterogeneous inverse demand elasticity case, we have

$$\begin{pmatrix} \widehat{\alpha} \\ \widehat{\beta} \end{pmatrix} = \left(I - \mathbb{E}\left[X_{it-1}X_{it}'\right]^{-1}\mathbb{E}\left[d_{iit}X_{it-1}X_{it}'\right]\right)\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$$

where X_{it} is the vector $(v_{it}, v_{it-1}^2)'$ and I is the identity matrix. It follows that the revenue markup satisfies

$$\widehat{\mu}_{it}^R = \left[1 - (\alpha + 2\beta\log V_{it})^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}'\left(\mathbb{E}\left[d_{iit}X_{it}X_{it-1}'\right]\mathbb{E}\left[X_{it}X_{it-1}'\right]^{-1}\right)\begin{pmatrix} 1 \\ 2\log V_{it} \end{pmatrix}\right](1 - d_{iit})^{-1}. \quad (1)$$

This markup is in general different from one for at least some firms. To see that clearly, let us further assume that the inverse demand elasticities are independent of the variable input usage and its square, such that, for any $n, m \in \mathbb{N}$, $\mathbb{E}\left[d_{iit}v_{it}^n v_{it-1}^m\right] = \mathbb{E}\left[d_{iit}\right]\mathbb{E}\left[v_{it}^n v_{it-1}^m\right]$.With these assumptions in place, one can show that $\widehat{\alpha} = \alpha(1 - \mathbb{E}\left[d_{iit}\right])$ and $\widehat{\beta} = \beta(1 - \mathbb{E}\left[d_{iit}\right])$. The revenue markup is equal to $\widehat{\mu}_{it}^R = (1 - \mathbb{E}\left[d_{iit}\right])(1 - d_{iit})^{-1}$ which is different from one since there exist a pair (i, j) such that $d_{iit} \neq d_{jjt}$. As for the Cobb-Douglas case, the bias is determined by an average inverse demand elasticities.

In the translog case, the average revenue markup is $\mathbb{E}\left[\log\widehat{\mu}_{it}^R\right] = \mathbb{E}\left[\log(\mu_{it})\right] + \mathbb{E}\left[\log\frac{\widehat{\alpha}+2\widehat{\beta}\log V_{it}}{\alpha+2\beta\log V_{it}}\right]$. Let us assume that the inverse demand elasticities are heterogeneous across firms in the sample. From equation (1), we can see that the average of the log revenue markup is equal to zero up to a Jensen like inequality:

$$\mathbb{E}\left[\log\widehat{\mu}_{it}^R\right] = -\mathbb{E}\left[\log(1 - d_{iit})\right] + \mathbb{E}\left[\log\left(1 - (\alpha + 2\beta\log V_{it})^{-1}\begin{pmatrix} \alpha \\ \beta \end{pmatrix}'\left(\mathbb{E}\left[d_{iit}X_{it}X_{it-1}'\right]\mathbb{E}\left[X_{it}X_{it-1}'\right]^{-1}\right)\begin{pmatrix} 1 \\ 2\log V_{it} \end{pmatrix}\right)\right].$$

When the inverse demand elasticities are homogeneous, $\forall i, d_{iit} = \gamma$, then the average log revenue markup is exactly zero.

The relationship between the average revenue and true markup now depends on the distribution of the variable input $\log V_{it}$ and the extent of the bias in the production function estimation. Importantly, the variance of the revenue markup is different from the variance of the true markup and also depends on the distribution of inputs and the covariance of input and the true markup. Finally, the correlation between the revenue and the true markup is no longer equal

[41] When $p_{it} = -\gamma y_{it}$, the vector $V^{-1}\begin{pmatrix} \mathbb{E}[p_{it}v_{it-1}] \\ \mathbb{E}[p_{it}v_{it-1}^2] \end{pmatrix} = \gamma\begin{pmatrix} \alpha \\ \beta \end{pmatrix}$ and the revenue markup becomes $\widehat{\mu}_{it}^R = (1 - \gamma)\frac{\alpha+2\beta\log V_{it}}{\alpha+2\beta\log V_{it}}(1-\gamma)^{-1} = 1$.

to one. To gauge the information contans of the revenue markup under translog production function, we rely on the simulation where we estimate true and revenue markup on simulated data.

C Derivation of the Translog Production Function

In this appendix we derive the translog approximation of the CES production function and show that it nests the Cobb Douglas production function. We specify a CES production function with homogeneity of degree γ:

$$Y_{iht} = \Omega_{iht} \left(\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} + (1-\alpha)[K_{iht}]^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}\gamma},$$

Cobb Douglas derivation The generalized CES production function nests the Cobb Douglas production function as $\eta \to 1$. To see this, note that:

$$\ln y_{iht} = \omega_{iht} + \frac{\eta}{\eta-1}\gamma \ln \left[\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} + (1-\alpha)[K_{iht}]^{\frac{\eta-1}{\eta}} \right].$$

The limit of this function for $\eta \to 1$:

$$\begin{aligned}
\lim_{\eta\to1} \ln y_{iht} &= \omega_{iht} + \gamma \lim_{\eta\to1} \frac{\ln \left[\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} + (1-\alpha)[K_{iht}]^{\frac{\eta-1}{\eta}} \right]}{\frac{\eta-1}{\eta}}, \\
&= \omega_{iht} + \gamma \lim_{\eta\to1} \frac{V_{iht}^{\frac{\eta-1}{\eta}} \ln V_{iht} \frac{\alpha}{\eta^2} + K_{iht}^{\frac{\eta-1}{\eta}} \ln K_{iht} \frac{(1-\alpha)}{\eta^2}}{1/\eta^2 \left(\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} + (1-\alpha)[K_{iht}]^{\frac{\eta-1}{\eta}} \right)}, \\
&= \omega_{iht} + \gamma \left[\ln V_{iht}\alpha + \ln K_{iht}(1-\alpha) \right],
\end{aligned}$$

where the second step follows from l'Hopital's rule. In levels, this yields the Cobb-Douglas production function with returns to scale γ:

$$Y_{iht} = \Omega_{iht} \left(V_{iht}^{\alpha} K_{iht}^{1-\alpha} \right)^{\gamma}.$$

Translog derivation The function implies the translog production function (11) up to a first order approximation around $\eta = 1$. To see this, start from:

$$\ln y_{iht} = \omega_{iht} + \frac{\eta}{\eta-1}\gamma \ln \left[\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} + (1-\alpha)[K_{iht}]^{\frac{\eta-1}{\eta}} \right],$$

$$\ln y_{iht} = \omega_{iht} + \frac{\eta}{\eta-1}\gamma \ln \left[\alpha [V_{iht}]^{\frac{\eta-1}{\eta}} \left(1 + \frac{(1-\alpha)}{\alpha} \left[\frac{K_{iht}}{V_{iht}} \right]^{\frac{\eta-1}{\eta}} \right) \right],$$

$$\ln y_{iht} = \omega_{iht} + \frac{\eta}{\eta - 1}\gamma \ln \left[\alpha[V_{iht}]^{\frac{\eta-1}{\eta}}\right] + \frac{\eta}{\eta - 1}\gamma \ln \left[1 + \frac{1-\alpha}{\alpha}\left(\frac{K_{iht}}{V_{iht}}\right)^{\frac{\eta-1}{\eta}}\right].$$

Moving the α back into the log term:

$$\ln y_{iht} = \omega_{iht} + \gamma v_{iht} + \frac{\eta}{\eta - 1}\gamma \ln \left[\alpha + (1-\alpha)\left(\frac{K_{iht}}{V_{iht}}\right)^{\frac{\eta-1}{\eta}}\right].$$

Consider the final term. Rewriting gives:

$$f(x) = \frac{\eta}{\eta - 1}\gamma \ln \left[1 + (1-\alpha)\left(\left(\frac{K_{iht}}{V_{iht}}\right)^{\frac{\eta-1}{\eta}} - 1\right)\right],$$

$$f(x) = \frac{\gamma}{x} \ln \left[1 + (1-\alpha)\left(B^x - 1\right)\right],$$

where $B = K_{iht}/L_{iht}$ and $x = (\eta - 1)/\eta$, such that our approximation is around $x \to 0$. Rewriting:

$$\begin{aligned} f(x) &= \frac{\gamma}{x} \ln \left[1 + (1-\alpha)\left((\exp(\text{x} \ln \text{B}) - 1\right)\right], \\ &\approx \frac{\gamma}{x} \ln \left[1 + (1-\alpha)\left(\text{x} \ln \text{B} - \frac{x^2[\ln B]^2}{2}\right)\right], \\ &\approx \frac{\gamma}{x} \left[(1-\alpha)\left(\text{x} \ln \text{B} - \frac{x^2[\ln B]^2}{2}\right) - \frac{(1-\alpha)^2}{2}\left(\text{x} \ln \text{B} - \frac{x^2[\ln B]^2}{2}\right)^2\right]. \end{aligned}$$

Given that we are approximating the function up to a first order we remove higher order terms, such that the final equation simplifies to:

$$f(x) = \frac{\gamma}{x}\left[(1-\alpha)\text{x} \ln \text{B} + \alpha\frac{1-\alpha}{2}x^2[\ln B]^2\right].$$

Hence the first order approximation of the generalized CES production function reads:

$$y_{iht} = \omega_{iht} + \gamma \ln V_{iht} + \gamma(1-\alpha) \ln \left(\frac{K_{iht}}{V_{iht}}\right) + \gamma\alpha\frac{1-\alpha}{2}\frac{\eta-1}{\eta}\left[\ln \left(\frac{K_{iht}}{V_{iht}}\right)\right]^2.$$

Grouping terms and denoting $x \equiv \ln X$:

$$y_{iht} = \omega_{iht} + \gamma\alpha v_{iht} + \gamma(1-\alpha)k_{iht} + \gamma\alpha\frac{1-\alpha}{2}\frac{\eta-1}{\eta}\left(v_{iht}^2 + k_{iht}^2 - 2k_{iht}v_{iht}\right),$$

which is the translog production function (11) with homogeneity of degree γ.

Variable input demand We next derive the demand for the variable input for the translog production function. The firms' cost minimization problem involves minimizing costs $W_t V_{iht}$

subject to the production function (11). Note that the output elasticity of the variable input is:

$$\frac{\partial y_{iht}}{\partial v_{iht}} = \gamma\alpha \left(1 + [1-\alpha]\frac{\eta-1}{\eta}\left[v_{iht} - k_{iht}\right]\right),$$

such that the first order condition of the cost minimization problem is:

$$W_t = \lambda_{iht}\frac{Y_{iht}}{V_{iht}}\gamma\alpha\left(1 + [1-\alpha]\frac{\eta-1}{\eta}\ln\left[\frac{V_{iht}}{K_{iht}}\right]\right)$$

where λ_{iht} is the Lagrange multiplier. Inverting the first order condition and inserting that marginal costs MC_{iht} equal λ_{iht}, we obtain (2).

Marginal costs As firms face an exogenous sequence of the fixed input K_{iht}, marginal costs can be derived from the production function (11) and optimal demand for the variable input (2). Inserting the latter into the former:

$$\begin{aligned} y_{iht} &= \omega_{iht} + \gamma\alpha\ln\left[\left(\frac{MC_{iht}}{W_t}\right)\gamma\alpha\left(1 - [1-\alpha]\frac{\eta-1}{\eta}\ln\left[\frac{K_{iht}}{V_{iht}}\right]\right)Y_{iht}\right] \\ &+ \gamma(1-\alpha)k_{iht} + \gamma\alpha\frac{1-\alpha}{2}\frac{\eta-1}{\eta}\left[\ln\left(\frac{K_{iht}}{V_{iht}}\right)\right]^2. \end{aligned}$$

Isolating log marginal costs on the left hand side yields (3).

For the calculation of the equilibrium, we use that firms within the sector are subject to the same sequence of factor prices. For a given price of a unit of the variable input W_t, the firm solves the static cost-minimization problem by choosing optimal variable input demand. This yields the first order condition:

$$V_{iht} = \left(\frac{MC_{iht}}{W_t}\right)\gamma\alpha\left(1 + [1-\alpha]\frac{\phi-1}{\phi}\ln\left[\frac{V_{iht}}{K_{iht}}\right]\right)Y_{iht}, \tag{2}$$

where, from inserting optimal variable demand into the production function, log marginal costs $mc_{iht} \equiv \ln MC_{iht}$ can be expressed as:

$$mc_{iht} = \ln\left[\frac{W_t}{\gamma}Y_{iht}^{\frac{1-\alpha\gamma}{\alpha\gamma}}\Omega_{iht}^{-\frac{1}{\alpha\gamma}}K_{iht}^{\frac{\alpha-1}{\alpha}}\right] - \ln\left(1 + [1-\alpha]\frac{\phi-1}{\phi}\ln\left[\frac{V_{iht}}{K_{iht}}\right]\right) + \frac{1-\alpha}{2}\frac{\phi-1}{\eta}\left(\ln\left[\frac{V_{iht}}{K_{iht}}\right]\right)^2 \tag{3}$$

D Additional Tables and Figures

Table A1: Estimation of AR(1) process for intermediate input prices

	(1)	(2)	(3)	(4)
Auto-regressive coefficient (ρ^w)	0.900***	0.871***	0.865***	0.868***
	(0.009)	(0.011)	(0.014)	(0.014)
St. Dev. of shocks (σ^w)	0.046	0.042	0.042	0.045
Controls	None	Year F.E.	Year F.E. & Ind. F.E.	Time Pol. & Ind. F.E.
Observations	798	798	798	798
R-squared	0.922	0.936	0.918	0.908

Note: Results from auto-regressions for intermediate input price indices (log) at the 2-digit level. Data from EU-KLEMS for France, 1995-2016. Standard errors in parentheses. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Time Pol. refers to the inclusion of a third-degree polynomial for time as a control.

Table A2: Estimation of AR(1) process for detrended nominal value added

	(1)	(2)	(3)	(4)
Auto-regressive coefficient (ρ^w)	0.999***	1.001***	0.677***	0.708***
	(0.00479)	(0.00412)	(0.0285)	(0.0257)
St. Dev. of shocks (σ^w)	0.166	0.140	0.419	0.390
Controls	None	Year F.E.	Year F.E. & Ind. F.E.	Time Pol. & Ind. F.E.
Observations	798	798	798	798
R-squared	0.982	0.987	0.709	0.608

Note: Results from auto-regressions for nominal sector-level value added (log) at the 2-digit level, detrended with nominal GDP. Data from EU-KLEMS for France, 1995-2016. Standard errors in parentheses. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Time Pol. refers to the inclusion of a third-degree polynomial for time as a control.

Table A3: Estimation of AR(1) process for fixed input using capital

	(1)	(2)	(3)	(4)
Auto-regressive coefficient (ρ^w)	0.988***	0.656***	0.656***	0.651***
	(0.000)	(0.008)	(008)	(0.002)
St. Dev. of shocks (σ^w)	0.215	0.215	0.662	11.79
Controls	None	Year F.E.	Year F.E. &	Ind-Year F.E. & Firm. F.E. Firm F.E.
Observations	160,124	160,124	160,124	160,124
R-squared	0.987	0.490	0.490	0.493

Note: Results from auto-regressions for French firms using EAP-FARE data for 2009-2018. Data on 27,857 firms. Standard errors in parentheses are clustered by firm. *, ** and *** denote statistical significance at the 10, 5 and 1% level, respectively. Industry fixed effects are at the 5-digit level.

Table A4: Change in Bias under Alternative Calibrations

Coefficients	True	Translog - Revenue (Full) $N_k = 16$	$N_k = 8$	$N_k = 4$	CD - Quantity (Full) $\phi = 1.05$	$\phi = 1.1$	$\phi = 1.3$
Production Function							
$\beta_v = \alpha\gamma$	0.32	0.77	0.67	0.67	0.32	0.31	0.29
		(0.027)	(0.0267)	(0.0213)	(0.003)	(0.003)	(0.002)
$\beta_k = (1-\alpha)\gamma$	0.48	0.14	0.21	0.18	0.49	0.49	0.52
		(0.0188)	(0.0159)	(0.011)	(0.001)	(0.001)	(0.001)
$\beta_{vv} = \gamma \frac{\alpha(1-\alpha)}{2} \frac{\phi-1}{\phi}$	0.009	0.05	0.034	0.043			
		(0.0053)	(0.0058)	(0.0047)			
$\beta_{kk} = \beta_{vv}$	0.009	0.013	0.005	0.003			
		(0.0025)	(0.0021)	(0.0015)			
$\beta_{vk} = -2\beta_{vv}$	-0.017	-0.057	-0.032	-0.038			
		(0.0068)	(0.0064)	(0.0043)			
AR(1) Productivity:							
ρ^ω	0.700	0.619	0.626	0.629	0.690	0.700	0.670
Implied parameters:							
α	0.400	0.849	0.762	0.792	0.39	0.39	0.36
γ	0.800	0.901	0.88	0.851	0.80	0.81	0.82
ϕ	1.100	1.977	1.251	1.377	1.00	1.00	1.00

Note: Estimated production function coefficients for different specifications. The top panel presents production function estimates. The bottom panel presents the deep parameters implied by the estimated production function.The first column presents true values for comparison. Bootstrapped standard errors are in parentheses.

Table A5: Overview - Productivity and Log Markup Estimates on different specifications

	Correlation Markup	Prod.	Error	Markup Moments (diff with true) Mean	St. Dev.	Median	IQR
True	1.00	1.00	1.00	0.00	0.00	0.00	0.00
Translog - Revenue (Full FS)							
Sector with $N_k = 16$	0.36	0.23	1.00	0.56	0.12	0.49	0.09
Sector with $N_k = 8$	0.77	0.23	1.00	0.54	0.05	0.50	-0.14
Sector with $N_k = 4$	0.77	0.21	1.00	0.51	0.06	0.46	-0.16
Cobb-Douglass - Quantity (Full FS)							
Sector with $\phi = 1.05$	0.93	1.00	1.00	0.01	0.02	0.02	0.06
Sector with $\phi = 1.1$	0.82	0.99	1.00	0.02	0.04	0.04	0.16
Sector with $\phi = 1.3$	0.57	0.94	1.00	0.04	0.18	0.10	0.66

Note: Table of moments of estimated productivity and markups. The first two columns present correlations of estimated productivity and markups with the true values. For *Revenue TL Full* and *Quantity CD Full* markup moments are reported as difference with the true ones, since every sector has different markup values.

Table A6: Correlations across Simulated Specifications - Log Markups

	True	Full - Q	Full - R	Basic - Q	Basic - R	None - Q	None - R
	Cobb-Douglas Production Function						
True	1.00	0.82	0.82	0.82	0.82	0.82	0.82
Full First Stage - Quantity	0.82	1.00	1.00	1.00	1.00	1.00	1.00
Full First Stage - Revenue	0.82	1.00	1.00	1.00	1.00	1.00	1.00
Basic First Stage - Quantity	0.82	1.00	1.00	1.00	1.00	1.00	1.00
Basic First Stage - Revenue	0.82	1.00	1.00	1.00	1.00	1.00	1.00
No First Stage - Quantity	0.82	1.00	1.00	1.00	1.00	1.00	1.00
No First Stage - Revenue	0.82	1.00	1.00	1.00	1.00	1.00	1.00
	Translog Production Function						
True	1.00	1.00	0.77	1.00	0.80	0.45	0.38
Full First Stage - Quantity	1.00	1.00	0.80	0.99	0.83	0.51	0.44
Full First Stage - Revenue	0.77	0.80	1.00	0.72	1.00	0.91	0.88
Basic First Stage - Quantity	1.00	0.99	0.72	1.00	0.75	0.38	0.32
Basic First Stage - Revenue	0.80	0.83	1.00	0.75	1.00	0.89	0.85
No First Stage - Quantity	0.45	0.51	0.91	0.38	0.89	1.00	0.99
No First Stage - Revenue	0.38	0.44	0.88	0.32	0.85	0.99	1.00

Note: Each cell presents the pairwise correlation between the markup in the row and the column header. The first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage. All markups are expressed in log. Simulated data.

Table A7: Correlations across Simulated Specifications - Log-Differenced Markups

	True	Full - Q	Full - R	Basic - Q	Basic - R	None - Q	None - R
	Cobb-Douglas Production Function						
True	1.00	0.89	0.89	0.89	0.89	0.89	0.89
Full First Stage - Quantity	0.89	1.00	1.00	1.00	1.00	1.00	1.00
Full First Stage - Revenue	0.89	1.00	1.00	1.00	1.00	1.00	1.00
Basic First Stage - Quantity	0.89	1.00	1.00	1.00	1.00	1.00	1.00
Basic First Stage - Revenue	0.89	1.00	1.00	1.00	1.00	1.00	1.00
No First Stage - Quantity	0.89	1.00	1.00	1.00	1.00	1.00	1.00
No First Stage - Revenue	0.89	1.00	1.00	1.00	1.00	1.00	1.00
	Translog Production Function						
True	1.00	1.00	0.85	1.00	0.87	0.55	0.44
Full First Stage - Quantity	1.00	1.00	0.87	0.99	0.89	0.59	0.48
Full First Stage - Revenue	0.85	0.87	1.00	0.83	1.00	0.89	0.83
Basic First Stage - Quantity	1.00	0.99	0.83	1.00	0.85	0.51	0.40
Basic First Stage - Revenue	0.87	0.89	1.00	0.85	1.00	0.87	0.81
No First Stage - Quantity	0.55	0.59	0.89	0.51	0.87	1.00	0.98
No First Stage - Revenue	0.44	0.48	0.83	0.40	0.81	0.98	1.00

Note: Each cell presents the pairwise correlation between the markup in the row and the column header. The first-stage regression includes a third-degree polynomial of inputs, price and market share controls, and time fixed effects ("Full"), only the polynomial and time fixed effects ("Basic") or no first stage. All markups are expressed in log. Simulated data.