Markups, Cost Shares, and Revenue Elasticities[*]

Stephen R. Bond
University of Oxford

Arshia Hashemi
NBER

January 13, 2020

Abstract

We explore the implications of the omitted price bias for the production approach to estimating markups. We distinguish between revenue and physical output elasticities. Under monopolistic competition and general demand and technology constraints, we show that if the revenue elasticity is used instead of the physical elasticity, then the production approach falsely implies a markup of one. We propose a new econometric methodology, the cost share approach, that is immune to the omitted price bias and that identifies firm heterogeneity in markups. We apply our cost share approach to UK micro data and find that exporters have positive markup premia.

[*]Bond: Nuffield College, University of Oxford, Oxford, UK (email: steve.bond@economics.ox.ac.uk). Hashemi: National Bureau of Economic Research, Cambridge, MA (e-mail: ahashemi@nber.org). This paper is a revised version of Chapter 1 of Hashemi's MPhil thesis at the University of Oxford (Hashemi, 2019). We thank seminar participants at the Harvard Graduate Workshop in IO and at the NBER for helpful comments and suggestions. This paper contains data from the UK Office for National Statistics (ONS) supplied by the UK Data Service (UKDS). The ONS and the UKDS bear no responsibility in relation to the either interpretation or further analysis of the data. All results have been reviewed to ensure that no confidential information is disclosed.

1 Introduction

A new literature on macroeconomic market power documents a rise in markups since the 1980s and relates this pattern to a number of secular trends, notably the fall in the labor share (De Loecker, Eeckhout, and Unger, 2019). This literature estimates markups using the production approach. Unlike the demand approach of Berry, Levinsohn, and Pakes (1995), the production approach imposes restrictions neither on the model of competition nor on the demand system. The essential assumption is that firms minimize costs. The firm's first order condition in cost minimization with respect to a flexible input identifies the firm markup as the ratio of the physical output elasticity of the input and its cost share in total revenue. The cost share is easily computed from production data, while the output elasticity is estimated using techniques in production function estimation. The production approach was pioneered by Hall (1988) and recently adapted for applications in microeconomic settings by De Loecker and Warzynski (2012).

A number of review articles including Basu (2019) and Syverson (2019) highlight open questions in this new field, such as the incongruencies between the micro estimates of the markups and the macro trends to be explained. We contribute to these recent evaluations by exploring the implications for the production approach of the omitted price bias highlighted in Klette and Griliches (1996) and more recently in Foster, Haltiwanger, and Syverson (2008). This bias arises when the empirical measure of output in the production function is total revenue divided by a sectoral price deflator. A consequence of the omitted price bias is that the coefficients on the factors in the estimating revenue production function are reduced form parameters that depend not only on the technology parameters of interest, but also on demand-side confounders. Thus, the estimated output elasticities are *revenue* output elasticities, rather than the physical output elasticities required for the production approach to work.

In an environment with monopolistic competition, a continuously differentiable production technology, and a downward sloping residual demand curve, we show that if the revenue output elasticity were used instead of the physical output elasticity, then the production approach falsely implies a markup of one for all firms and time periods. This is the central theoretical result of this paper. We propose an alternative econometric methodology that identifies firm heterogeneity in markups in the absence of data on output prices and physical output quantities. This is the cost share approach that does not rely on production function estimation. Under our cost share approach, the first order condition in cost minimization is used to construct a regression equation in which the dependent variable is the logarithm of the cost share and the explanatory variables are the logarithm of the physical output elasticity and the logarithm of the markup. Prior to estimation, the researcher must impose functional form assumptions for the physical output elasticity and the markup. Given these assumptions, the parameters pertaining to firm heterogeneity in markups are identified and can be estimated consistently using standard regression or method of moments techniques under appropriate orthogonality conditions.

We apply our cost share approach to investigate whether there are markup premia associated with exporting and importing using UK micro data on production and trade for the ten year period 2003-2012. In line with the results of De Loecker and Warzynski (2012), we find that exporters have positive, statistically significant, and economically meaningful markup premia. The positive markup premia of exporters are robust to the choice of functional forms assumptions for the cost share regression. The evidence on the markup premia of importers is mixed. We emphasize that we do not interpret these estimated markup premia in a causal manner.

The remainder of this paper is organized as follows. Section 2 outlines the production approach to markups. Section 3 explores the implications of the omitted price bias for the production approach. Section 4 details our cost share approach. Section 5 describes the micro data and presents our empirical results. Section 6 concludes.

2 Production approach to markups

This section outlines the production to approach to estimating markups. This framework derives an expression for the markup without specifying a model of competition, a demand system, or a functional form for the production technology. The essential assumption is behavioral, namely that firms minimize production costs.

2.1 Cost minimization

Consider an economy in which firms are heterogeneous with respect to their productivity Ω_{it} and produce physical output Q_{it} using the following production technology

$$Q_{it} = Q\left(\Omega_{it}, \boldsymbol{V_{it}}, K_{it}; \boldsymbol{\theta}\right) \tag{1}$$

where $\boldsymbol{V}_{it} = \left(V_{it}^{1}, ..., V_{it}^{J}\right)'$ is a vector of J flexible factors of production, K_{it} is the capital stock, and $\boldsymbol{\theta}$ is a vector of technology parameters. An important statistic is the physical output elasticity.

Definition 1. The physical output elasticity of a flexible input V_{it}^v is defined as

$$\theta_{it}^v \equiv \frac{V_{it}^v}{Q_{it}} \frac{\partial Q_{it}}{\partial V_{it}^v}$$

In period t, the firm's choice variables are the flexible inputs in $\boldsymbol{V}_{it}$ and its state variables are (K_{it}, Ω_{it}). The production approach makes a very general assumption about the conduct of firms in the output and input markets.

Assumption 1. *Firms are price setters in the output market and price takers in the input markets.*

Let $\boldsymbol{W_t} = \left(W_t^1, ..., W_t^J\right)'$ denote the vector of input prices for the flexible inputs and let R_t denote the unit price of capital. Given price taking behavior in the input markets, firms take the input prices $\boldsymbol{W_t}$ and R_t as given. In each period t, the firm minimizes its total production costs subject to satisfying an output constraint. The Lagrangian function for this constrained optimization problem is

$$\mathcal{L}\left(\boldsymbol{V_{it}}, K_{it}, \lambda_{it}\right) = \sum_{v=1}^{J} W_t^v V_{it}^v + R_t K_{it} + \lambda_{it}\left(\bar{Q}_{it} - Q_{it}\left(\cdot\right)\right)$$

The Lagrange multiplier λ_{it} is the shadow value of the output constraint and describes the change in the value of the objective function as we relax the output constraint. The first order condition with respect to the flexible input V_{it}^v is

$$W_t^v = \lambda_{it} \frac{\partial Q_{it}}{\partial V_{it}^v}$$

Let P_{it} denote the output price and note that λ_{it} is a measure of marginal cost. We define the markup as the ratio of the output price to marginal cost, $\mu_{it} \equiv P_{it}/\lambda_{it}$. Substituting for λ_{it} in the first order condition and rearranging yields an expression for the markup

$$\mu_{it} = \frac{\theta_{it}^v}{\alpha_{it}^v} \tag{2}$$

where $\alpha_{it}^v = \left(W_t^v V_{it}^v\right) / \left(P_{it} Q_{it}\right)$ is the cost share of the input V_{it}^v in total revenue. Equation (2) identifies the markup as a function of two objects: the physical output elasticity θ_{it}^v; and the cost share α_{it}^v. While the cost share can be computed directly using data on total revenue and expenditure on inputs, the physical output elasticity is unobservable and must be estimated.

2.2 Production function estimation

The primary method for estimating θ_{it}^{v} is production function estimation. A central issue in estimating production functions is the simultaneity bias (Marschak and Andrews, 1944). Firms choose their factors of production optimally in response to productivity shocks, which are observable to the firm, but not to the econometrician. The standard practice for addressing this endogeneity problem is the proxy approach of Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg, Caves, and Frazer (2015). We sketch the estimation procedure by focusing on a two factor Cobb-Douglas production function in labor L_{it} and capital K_{it} and using intermediate inputs M_{it} as the proxy for unobserved productivity $\omega_{it} = \ln\Omega_{it}$. We assume that quantity data on $\{Q_{it}, L_{it}, K_{it}, M_{it}\}$ are available for a panel of firms. The estimating production function is

$$q_{it} = \theta^{L} l_{it} + \theta^{K} k_{it} + \omega_{it} + \epsilon_{it} \tag{3}$$

The lower case variables denote logarithms of the original variables and $\theta = \left(\theta^{L}, \theta^{K}\right)'$ is the parameter vector of interest. Note that the physical output elasticities in θ are constant and therefore common to all firms and time invariant. The implication is that all firm and time variation in the markup is driven solely by variation in the cost share. The productivity shock ω_{it} is potentially correlated with the factors (l_{it}, k_{it}). Meanwhile, ϵ_{it} captures measurement error in output and ex post shocks to production and is thus assumed to be uncorrelated with the factors. The necessary assumptions for ensuring that m_{it} is a valid proxy for productivity are scalar unobservability and strict monotonicity.

Assumption 2. *Firm productivity is the only scalar unobservable entering the input demand equation for intermediate inputs and the demand for intermediate input is strictly monotonic in productivity.*

$$m_{it} = g_t\left(l_{it}, k_{it}, \omega_{it}\right) \tag{4}$$

Inverting the input demand equation in (4) yields an expression for unobserved productivity as a function of observable variables.

$$\omega_{it} = h_t\left(l_{it}, k_{it}, m_{it}\right) \tag{5}$$

The right-hand side of equation (5) is the proxy for firm productivity. It is standard practice to assume that the law of motion for productivity follows a first order Markov process.

$$\omega_{it} = g\left(\omega_{i,t-1}\right) + \xi_{it} \tag{6}$$

The term ξ_{it} is the serially uncorrelated innovation to productivity. Ackerberg, Caves, and Frazer (2015) propose a two stage estimation procedure. In the first stage, we use OLS to estimate the model

$$q_{it} = \phi_t\left(l_{it}, k_{it}, m_{it}\right) + \epsilon_{it}$$

where $\phi_t\left(l_{it}, k_{it}, m_{it}\right) = \theta_l l_{it} + \theta_k k_{it} + h_t\left(l_{it}, k_{it}, m_{it}\right)$ is expected output.[1] The first stage produces estimates of expected output $\hat{\phi}_{it}$. Using the law of motion in equation (6), the innovation to productivity can be recovered as a function of θ and the data.

$$\xi_{it}\left(\theta\right) = \left(\hat{\phi}_{it} - \theta^L l_{it} - \theta^K k_{it}\right) - g\left(\hat{\phi}_{i,t-1} - \theta^L l_{i,t-1} - \theta^K k_{i,t-1}\right)$$

In the second stage, we estimate the parameters of interest using the following moment conditions.

$$\mathbb{E}\left[\xi_{it}\left(\theta\right)\begin{pmatrix} l_{i,t-1} \\ k_{it} \end{pmatrix}\right] = 0 \tag{7}$$

These moment conditions rely on timing assumptions, namely that labor is a flexible input that responds contemporaneously to the productivity innovation, whereas capital is dynamic input determined one period in advance.

3 Omitted price bias

Researchers rarely observe separate price and quantity data. A standard measure of physical output in the production function is total revenue divided by a sectoral price deflator. Klette and Griliches (1996) explore the implications of this so-called omitted price bias for production function estimation. Suppose that the researcher observes total revenue $S_{it} = P_{it}Q_{it}$ rather than physical output Q_{it}. Denote by P_t the observable industry output price deflator and by $\tilde{S}_{it} = P_{it}Q_{it}/P_t$ deflated revenue.We define the revenue output elasticity as follows.

[1] The proxy variable $h_t\left(l_{it}, k_{it}, m_{it}\right)$ is treated nonparametrically using a high order polynomial in its arguments.

Definition 2. The revenue output elasticity of a flexible input V_{it}^{v} is defined as

$$\gamma_{it}^{v} \equiv \frac{V_{it}^{v}}{\tilde{S}_{it}} \frac{\partial \tilde{S}_{it}}{\partial V_{it}^{v}}$$

Without loss of generality, we concentrate on a Cobb-Douglas technology in labor and capital. The revenue production function is

$$\tilde{s}_{it} = \theta^{L} l_{it} + \theta^{K} k_{it} + \omega_{it} + (p_{it} - p_{t}) \tag{8}$$

The new error term $(p_{it} - p_t)$ is the omitted price variable, which is non-zero when firms have price setting power in the output market. Market power is one reason why the omitted price variable may be negatively correlated with the factors of production. Firms with market power charge prices above the industry average, produce less output, and therefore purchase fewer inputs. Operating under general conditions, Klette and Griliches show that the omitted price bias induces a downward bias in the estimates of returns to scale.

3.1 Klette and Griliches

We extend the work of Klette and Griliches to study the implications of the omitted price bias for the production approach to estimating markups. The Klette and Griliches economic environment comprises monopolistic competition, a Cobb-Douglas production function, and a CES demand system.

$$Q_{it}\left(\Omega_{it}, L_{it}, K_{it}\right) = \Omega_{it} L_{it}^{\theta^{L}} K_{it}^{\theta^{K}} \tag{9}$$

$$Q_{it}\left(P_{it}\right) = Q_{t}\left(\frac{P_{it}}{P_{t}}\right)^{-\sigma} \tag{10}$$

Under monopolistic competition, the atomistic nature of firms means that firms take the sectoral output Q_t and the sectoral price level P_t as given. The price elasticity of demand is equal to the constant elasticity of substitution parameter σ. We use the demand system to solve for the omitted price variable.

$$p_{it} - p_{t} = \frac{q_{t} - q_{it}}{\sigma}$$

Given the functional forms in (9) and (10), we use substitute for the omitted price variable in equation (8) to express the revenue production function in terms of the revenue output elasticities.

$$\tilde{s}_{it} = \gamma^L l_{it} + \gamma^K k_{it} + \tilde{\omega}_{it} + \tilde{q}_t$$

where the revenue output elasticities are given by

$$\gamma^L = \left(\frac{\sigma - 1}{\sigma}\right)\theta^L$$
$$\gamma^K = \left(\frac{\sigma - 1}{\sigma}\right)\theta^K$$

and $\tilde{q}_t$ is a time fixed effect. It is apparent that the revenue output elasticity of labor γ^L is strictly less than the physical output elasticity θ^L when the demand elasticity σ is finite. In the limiting case of perfect competition, we have $\sigma \to \infty$, and the revenue and physical output elasticities coincide exactly because the omitted price variable equals zero. It is straightforward to estimate the revenue output elasticities $\gamma = \left(\gamma^L, \gamma^K\right)'$ using exactly the same production function estimation techniques outlined in Section 2.2, except with access to data on $\left\{\tilde{S}_{it}, L_{it}, K_{it}, M_{it}\right\}$ rather than on $\{Q_{it}, L_{it}, K_{it}, M_{it}\}$. The following lemma is helpful for proving our first theoretical result.

Lemma 1. *Under monopolistic competition and a CES demand system, the markup can be expressed as a function of only the price elasticity of demand.*

$$\mu_{it} = \frac{\sigma}{\sigma - 1}$$

Proof. See Appendix A.1

□

The combination of monopolistic competition and a CES demand system implies constant markups across all firms and time periods (De Loecker, 2011). We prove that if the revenue elasticity of labor were used instead of its physical output elasticity for the purposes of estimating markups under the production approach, then the downward bias from omitted prices is so severe that the production falsely implies a markup of one for all firms and time periods.

Proposition 1. *Under monopolistic competition, a Cobb-Douglas production function, and a CES demand system, if the revenue output elasticity of labor γ^L were used in place of the* *physical output elasticity θ^L in equation (2), then we have*

$$\frac{\gamma^L}{\alpha_{it}^L} = 1$$

Proof. The Cobb-Douglas technology and the CES demand system imply that we can express the revenue elasticity γ^L as a function of the demand elasticity σ and the physical output elasticity θ^L

$$\gamma^L = \left(\frac{\sigma - 1}{\sigma}\right)\theta^L$$

Lemma 1 shows that, under monopolistic competition, the markup μ_{it} can be expressed as a function of the demand elasticity σ

$$\mu_{it} = \frac{\sigma}{\sigma - 1}$$

Lastly, cost minimization implies that

$$\mu_{it} = \frac{\theta^L}{\alpha_{it}^L}$$

Combining everything together yields the desired result

$$\begin{aligned}\frac{\gamma^L}{\alpha_{it}^L} &= \left(\frac{\sigma - 1}{\sigma}\right)\frac{\theta^L}{\alpha_{it}^L} \\ &= \frac{1}{\mu_{it}}\mu_{it} \\ &= 1\end{aligned}$$

□

3.2 Generalization

The drawback of the Klette and Griliches framework is that its CES demand assumption implies constant markups under monopolistic competition. In order to generate endogenously variable markups, one can either deviate from monopolistic competition or from CES demand. We retain the assumption of monopolistic competition and generalize the demand system to any downward sloping residual demand curve, which nests the isoelastic CES demand function and linear demand (Melitz and Ottaviano, 2008). In addition, we do not impose a functional form for the production function; the only restriction we impose is that the production technology is continuous and differentiable. Our next proposition proves that the result of Proposition 1 still applies in this more general environment.

Proposition 2. *Under monopolistic competition, a continuously differentiable production technology $Q_{it} = Q\left(\Omega_{it}, \boldsymbol{V_{it}}, K_{it}; \boldsymbol{\theta}\right)$ and a downward sloping demand system $Q_{it}\left(P_{it}\right)$, if the revenue output elasticity of a flexible input, γ_{it}^{v}, were used instead of the physical output elasticity, θ_{it}^{v}, in equation (2), then we have*

$$\frac{\gamma_{it}^{v}}{\alpha_{it}^{v}} = 1, \forall v \in \{1, ..., J\}$$

Proof. See Appendix A.2.

□

Unlike the constant revenue output elasticity in Proposition 1, the revenue elasticity in Proposition 2 may well vary across firms and time due to variation in either the demand elasticity or the physical output elasticity or both. Proposition 2 shows that the production approach falsely implies that firms have no market power when in fact monopolistic firms face an inelastic demand curve and hence have price setting power in the output market.

4 Cost share approach

We propose a simpler method for studying firm heterogeneity in markups. This is our novel cost share approach that does not require consistent estimates of physical output elasticities. In fact, the cost share approach does not rely on production function estimation at all. To illustrate, consider the expression for markup in equation (2). Suppose that there is additive measurement error in the observed log cost share, $\ln\tilde{\alpha}_{it}^{v} = \ln\alpha_{it}^{v} + \epsilon_{it}$, where α_{it}^{v} is the latent cost share and the measurement error ϵ_{it} is mean zero and serially uncorrelated. Then, taking logs of equation (2) gives the general form of our cost share regression equation.

$$\ln\tilde{\alpha}_{it}^{v} = \ln\theta_{it}^{v} - \ln\mu_{it} + \epsilon_{it} \tag{11}$$

Before estimating this regression model, the researcher must make two functional form assumptions. The first is the choice of functional form for the production technology that determines the functional form of the physical output elasticity θ^v_{it}. The second is the choice of functional form for the markup process $\ln\mu_{it}$. We turn to a discussion of possible functional form choices in the context of studying firm heterogeneity in markups between exporters and non-exporters. Throughout, we assume the production function is Hicks-neutral in productivity.

4.1 Cobb-Douglas

We follow De Loecker and Warzynski (2012) in expressing the log markup as a linear function of an export status dummy variable denoted by e_{it}, which equals one if firm i has positive exporting revenue at time t and zero otherwise.

$$\ln\mu_{it} = \gamma_0 + \gamma_1 e_{it} + \mathbf{b}'_{\mathbf{it}}\boldsymbol{\delta} + \nu_{it} \tag{12}$$

The vector $\mathbf{b}_{it}$ contains a set of control variables such as time fixed effects and employment. Suppose further that the production technology is Cobb-Douglas so that the physical output elasticity of a variable input v is a constant denoted by θ^v. Our Cobb-Douglas cost share regression equation is

$$\ln\tilde{\alpha}^x_{it} = \lambda_0 + \lambda_1 e_{it} + \mathbf{b}'_{\mathbf{it}}\boldsymbol{\phi} + \varepsilon_{it} \tag{13}$$

where $\lambda_0 = (\ln\theta^v - \gamma_0)$, $\lambda_1 = -\gamma_1$, $\boldsymbol{\phi} = -\boldsymbol{\delta}$, and $\varepsilon_{it} = (\epsilon_{it} - \nu_{it})$. The physical output elasticity $\ln\theta^v$ is not separately identified from the intercept term γ_0 in the markup process. However, the slope coefficient γ_1 that describes the markup premia of exporters in percentage terms is identified and given by $-\lambda_1$. If export status e_{it} is orthogonal to the error term in the markup equation, $\mathbb{E}\left(e_{it}\nu_{it}\right) = 0$, and also orthogonal to measurement error in the cost share, $\mathbb{E}\left(e_{it}\epsilon_{it}\right) = 0$, then the OLS estimator of the parameter λ_1 (and hence of γ_1) is consistent. Crucially, the markup premia of exporters is identified and estimated consistently using OLS without relying on production function estimation to estimate θ^v and other physical output elasticities.

4.2 Translog

The cost share approach can be generalized to incorporate the more flexible translog production technology. We continue to retain the markup process in equation (12). Without loss of generality, we consider a three-factor translog production function in labor l_{it}, intermediate inputs m_{it}, and capital k_{it}.

$$q_{it} = \theta^L l_{it} + \theta^{LL} l_{it}^2 + \theta^K k_{it} + \theta^{KK} k_{it}^2 + \theta^M m_{it} + \theta^{MM} m_{it}^2 + \theta^{KL} k_{it} l_{it} + \theta^{KM} k_{it} m_{it} + \theta^{LM} l_{it} m_{it} + \omega_{it}$$

We consider m_{it} as the flexible input. Its physical output elasticity is

$$\frac{\partial q_{it}}{\partial m_{it}} = \theta^M + 2\theta^{MM} m_{it} + \theta^{KM} k_{it} + \theta^{LM} l_{it}$$

Note that the physical output elasticity under a translog technology varies across both firms and time due to variation in the use of factors across firms and time. Our translog cost share regression equation is

$$\ln\tilde{\alpha}_{it}^M = \ln\left(\theta^M + 2\theta^{MM} m_{it} + \theta^{KM} k_{it} + \theta^{LM} l_{it}\right) + \lambda_0 + \lambda_1 e_{it} + \mathbf{b}_{\mathbf{it}}' \phi + \varepsilon_{it} \tag{14}$$

where $\lambda_0 = -\gamma_0$, and λ_1, $\boldsymbol{\phi}$, and ε_{it} are defined as in the Cobb-Douglas case. The markup parameters λ_0, λ_1, and $\boldsymbol{\phi}$ enter equation (14) linearly, whereas the technology parameters θ^M, θ^{MM}, θ^{KM}, and θ^{LM} enter nonlinearly. Under appropriate orthogonality conditions, both sets of parameters can be estimated consistently using either nonlinear least squares or nonlinear GMM.[2] We note that in this partially linear model, the control variables in the vector $\mathbf{b}_{\mathbf{it}}$ must not be collinear with the factors of production in the logarithm function. Thus, a consequence of incorporating the more flexible translog production technology is that greater restrictions must be imposed on the markup process to avoid issues of collinearity.

4.3 Nonparametric

In the most general case, the researcher is agnostic about the choice of production technology and treats the production function nonparametrically. In the case of a three factor production function, the physical output elasticity of intermediate inputs is some unknown function of all three factor inputs, $\partial q_{it}/\partial m_{it} = f\left(l_{it}, m_{it}, k_{it}\right)$. We continue to retain the markup process in equation (12). Our nonparametric cost share regression equation is

$$\ln\tilde{\alpha}_{it}^M = g\left(l_{it}, m_{it}, k_{it}\right) + \lambda_1 e_{it} + \mathbf{b}_{\mathbf{it}}' \phi + \varepsilon_{it} \tag{15}$$

where $g\left(l_{it}, m_{it}, k_{it}\right) = \ln f\left(l_{it}, m_{it}, k_{it}\right) - \gamma_0$ and λ_1, $\boldsymbol{\phi}$, and ε_{it} are defined as in the Cobb-Douglas case. The key parameter of interest is the slope coefficient λ_1 identifying the markup

[2]Since m_{it} appears as an explanatory variable in equation (14), we require that the source of the measurement error in the cost share be from revenue ($P_{it}Q_{it}$) rather than from total expenditure on intermediate inputs ($W_{it}^M M_{it}$).

premia of exporters, and $g(l_{it}, m_{it}, k_{it})$ is an incidental or nuisance function that must also be estimated. Since the function $g(l_{it}, m_{it}, k_{it})$ is treated nonparametrically, the intercept term $-\gamma_0$ is subsumed in this function and is therefore not identified. The researcher can estimate equation (15) either by approximating the unknown function $g(l_{it}, m_{it}, k_{it})$ using a high order polynomial in its arguments or by using semiparametric methods for partially linear models (Robinson, 1988). Like the translog case, the nonparametric case requires that the control variables in the vector $\mathbf{b_{it}}$ not be collinear with the factors in the function $g(l_{it}, m_{it}, k_{it})$.

4.4 Other extensions

Our cost share approach can be extended in numerous other ways. For example, the markup process could depend on continuous explanatory variables as well as discrete ones such as export status. We could allow for endogenous explanatory variables in the markup process if valid and relevant instrumental variables are available. If the estimating sample pools observations across different sectors, then we can allow for sector specific physical output elasticities by including sector dummies rather than a single intercept term in equation (13). Moreover, we can allow for time-invariant firm heterogeneity in the physical output elasticities by considering firm fixed effects rather than sector dummies. If we are interested in trends in markups over time, as in De Loecker, Eeckhout, and Unger (2019), we could investigate this by including trend terms or time dummies as explanatory variables in the markup process. Finally, if we are interested in the persistence of markups, we could explore this by specifying a dynamic process for the explanatory variables in the markup process and using suitable dynamic panel data methods to estimate the resulting specifications (Arellano and Bond, 1991). We leave these extensions open for future research.

5 Data and results

We apply our cost share approach to study firm heterogeneity in markups for exporting and importing firms. In Sections 5.1 and 5.2, we describe our micro data and document descriptive statistics. In Section 5.3, we report the main empirical results.

5.1 Data

The micro data used in estimation is provided by the UK's Office for National Statistics (ONS). Our primary data source is the Annual Respondents Database X (ARDx, ONS, 2017), which comprises yearly firm level surveys. We append the yearly surveys to construct a panel data set for the ten year period 2003-2012. Three features of the ARDx make it ideal for studying firm heterogeneity in markups. First, the data are reported at the "reporting unit" level at which production activity takes place. A reporting unit could be either a standalone firm or part of an "enterprise", which comprises a collection of reporting units that jointly constitute a firm. This feature makes it possible to evaluate the cost shares at the firm level. Second, the data include empirical measures with which to construct the cost shares. These are total revenue, yearly employment, and expenditure on intermediate inputs. Our measure of the capital stock is sourced from the researcher constructed data set of Trushin et al. (2015), who apply the perpetual inventory method to annual ONS investment data. Where appropriate, we use the ONS sectoral output and input price deflators to deflate revenue, expenditure on intermediate inputs, and the capital stock. Third, the data include indicators of international trade activity such as exporting revenue and expenditure on imports.

Our secondary data source is the Business Expenditure on Research and Development (BERD, ONS, 2019), which reports total expenditure on R&D. We merge the BERD data with the ARDx data using the unique firm identifier. The BERD data allow us to identify the firms in the ARDx that are active in R&D. It is important to distinguish between R&D and non-R&D firms when estimating the cost share regressions. The reason is that the cost share of R&D firms confounds factor contributions to production with factor contributions to R&D, whereas our empirical framework requires that the cost share measure capture only factor contributions to production. The BERD data do not separately report input expenditures on production and on R&D. Therefore, we perform a robustness check by estimating our regressions on separate samples that include and exclude R&D firms. We impose three restrictions in constructing the base sample for empirical analysis. First, we focus on firms in the manufacturing industry (two-digit SIC codes 10 to 32). Second, we exclude firms with fewer than ten employees. Third, we drop firms for which we have observations on two consecutive years or fewer. These restrictions yield a sample in which there are approximately 5,000 reporting units, of which approximately 4,700 are single establishment firms.

5.2 Descriptive statistics

We document the frequency of exporting, importing, and R&D activity and the performances of exporting, importing, and R&D firms across a number of outcomes measures. Table 1 presents the breakdown of firms in the ARDx by two-digit manufacturing sector and classifies the number of firms that are exporters, importers, and R&D investors. Specifically, we label a firm as an exporter if it has exported at least once in the ten year period 2003-2012. The definitions for importers and R&D investors are analogous.

Table 1: Exporting, Importing, and R&D Firms by Two-Digit Manufacturing Sector

Manufacturing sector	Firms	Export	%	Import	%	R&D	%
Food, beverages, tobacco	868	176	20.3	313	36.1	457	52.6
Textiles	261	53	20.3	104	39.8	117	44.8
Wearing apparel	103	15	14.6	26	25.2	32	31.1
Paper	240	64	26.7	104	43.3	95	39.6
Printing and recording	203	65	32.0	71	35.0	48	23.6
Refined petroleum	33	11	33.3	12	36.4	16	48.5
Chemicals	419	182	43.4	253	60.4	314	74.9
Pharmaceuticals	80	51	63.8	58	72.5	70	87.5
Rubber and plastic	308	97	31.5	135	43.8	163	52.9
Basic metals	283	88	31.1	146	51.6	107	37.8
Fabricated metals	530	165	31.1	204	38.5	251	47.4
Computers and electronics	401	158	39.4	212	52.9	310	77.3
Electrical equipment	537	231	43.0	288	53.6	358	66.7
Machinery and equipment	269	102	37.9	149	55.4	177	65.8
Motor vehicles	142	62	43.7	77	54.2	90	63.4
Other transport equipment	153	29	19.0	46	30.1	77	50.3
Furniture	249	92	36.9	128	51.4	202	81.1
Aggregate manufacturing	439	132	32.2	194	45.9	250	56.9
Observations	24,222						

Note: The aggregate manufacturing row presents the average values from pooling observations across sectors.

The empirical literature on micro level trade patterns documents that exporting is rare (Bernard and Jensen, 1995). Table 1 confirms the typical empirical finding that only some firms within each sector are exporters. For the UK manufacturing industry as a whole, approximately 32 percent of firms export. Moreover, the fraction of exporters varies substantially by sector, from around 64 percent in pharmaceuticals to around 15 percent in wearing apparel. The fraction of importers is 46 percent, which is somewhat higher than the fraction of exporters. The fraction of R&D investors is the highest; an average of 57 percent of all manufacturing firms are R&D investors. The reason why our sample reports a high intensity of R&D operations is that our empirical measure of R&D is a binary indicator for whether a firm has planned R&D investments, which is an input of the R&D process. This contrasts with alternative measures of R&D in related studies, notably Aghion et al. (2018), who use data on the number of patents issued, which is an output of the R&D process. The BERD data report neither the number of patents issued nor on any other output of R&D.

Table 2: Exporter, Importer, and R&D Premia

	(1) Sales	(2) Employment	(3) VA/Worker	(4) Wage/Worker	(5) Capital/Worker
Exporter	0.07***	0.15***	0.12***	0.10***	0.05*
	(4.72)	(6.36)	(8.30)	(12.97)	(2.55)
Importer	0.22***	0.37***	0.21***	0.08***	0.24***
	(18.55)	(19.20)	(18.08)	(13.64)	(14.85)
R&D investor	-0.03*	0.49***	0.01	0.05***	-0.01
	(-2.52)	(29.16)	(0.50)	(9.94)	(-0.73)
Observations	21,427	21,427	21,427	21,427	21,427

t statistics in parentheses. Standard errors are clustered at the firm-level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2 presents evidence on the performance of exporters, importers, and R&D investors across a number of outcomes. Our measure of labor productivity is value added per worker. In each column, we regress the log of the outcome variable on binary indicators for whether a firm is an exporter, an importer, and an R&D investor. We include full sector and year fixed effects as controls in each regression, and all regressions except that in column 2 include log employment as a control for firm size. As in the literature on heterogeneous firms following Melitz (2003), the self-selection of firms into exporting and R&D implies systematic differences in the performance measures between firms that select into these activities and those that do not. The UK micro data show that there are sizeable and statistically significant premia associated with trade and R&D activity. Controlling for import and R&D status, firms that export are larger in terms of sales and employment, have higher labor productivity, pay higher wages, and are more capital intensive. Importing firms exhibit similar premia along these dimensions. Firms that invest in R&D are larger in terms of employment and pay higher wages.

5.3 Empirical results

In all of our cost share regressions, we consider intermediate inputs m_{it} as the flexible input with respect to which the first order condition in cost minimization is taken. Therefore, the dependent variable in our cost share regressions is the log cost share of intermediate inputs in total revenue. We estimate two types of cost share regressions: a linear model and a nonlinear model.[3] We report results for two linear specifications. The first is based on the Cobb-Douglas production technology, which implies that the physical output elasticity is constant for all firms in the same sector. We control for this output elasticity using sector fixed effects. The second is for the nonparametric technology. We approximate the incidental function of the factors in equation (15) using a third order polynomial. Our nonlinear model pertains to the translog production technology and we estimate this model using nonlinear least squares.

Table 3: Cost Share Regression: Cobb Douglas Technology

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Export status	0.029	0.063	-0.011	0.007				
	(1.22)	(1.70)	(-0.50)	(0.21)				
Both					-0.028	-0.033	-0.074**	-0.087
					(-0.97)	(-0.71)	(-2.65)	(-1.96)
Exporter only					0.081*	0.157**	0.047	0.094
					(2.06)	(2.88)	(1.28)	(1.79)
Importer only					-0.135***	-0.197***	-0.122***	-0.171***
					(-6.20)	(-7.41)	(-6.03)	(-6.97)
Fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R&D firms	Yes	No	Yes	No	Yes	No	Yes	No
Observations	24,222	12,170	24,222	12,170	24,222	12,170	24,222	12,170
R^2	0.000	0.001	0.066	0.077	0.004	0.009	0.070	0.083

t statistics in parentheses. Standard errors are clustered at the firm-level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 presents the results for the linear cost share regression in which the production technology is Cobb-Douglas. Columns 1-4 of Table present results for a single markup shifter, which is firm export status. The fixed effects indicator refers to whether the model includes

[3]We categorize a regression model as linear if it is linear in parameters and as nonlinear if at least some of its parameters appear nonlinearly.

full sector and year fixed effects. The R&D firms indicator refers to whether the regression is estimated on the sample that includes R&D firms. Columns 5-8 present results for a series of binary indicators that classify whether a firm is an exporter, an importer, or both. The reference group is the set of firms that are neither exporters nor importers. In the single regressor case, there is no evidence of statistically significant markup premia for exporters. In the multiple regressor case, there is evidence of large and positive markup premia for firms that are exporters only, although these premia are only statistically significant if fixed effects are excluded from the model. Moreover, there is evidence of negative and significant markup premia for firms that are importers only.

Table 4: Cost Share Regression: Translog Technology

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Export status	0.076***	0.101***	0.031	0.053				
	(3.85)	(3.37)	(1.58)	(1.81)				
Both					0.049*	0.061	0.015	0.010
					(2.03)	(1.58)	(0.64)	(0.26)
Exporter only					0.079*	0.128**	0.038	0.082
					(2.48)	(2.86)	(1.23)	(1.88)
Importer only					-0.067***	-0.103***	-0.039*	-0.094***
					(-3.68)	(-4.73)	(-2.21)	(-4.76)
Fixed effects	No	No	Yes	Yes	No	No	Yes	Yes
R&D firms	Yes	No	Yes	No	Yes	No	Yes	No
Observations	24,222	12,170	24,222	12,170	24,222	12,170	24,222	12,170
R^2	0.319	0.334	0.352	0.368	0.322	0.336	0.352	0.370

t statistics in parentheses. Standard errors are clustered at the firm-level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 4 presents the results for the nonlinear cost share regressions in which the production technology is translog. In this model, the production function parameters enter nonlinearly, while the markup parameters appear linearly, as in equation (14). We estimate this model using nonlinear least squares. Without fixed effects, there are positive, statistically significant, and economically meaningful markup premia for firms that are exporters only. When fixed effects are included, these premia are smaller and become insignificant. Like the Cobb-Douglas case, the translog case provides evidence of negative and statistically significant markup premia for importers.

Table 5: Cost Share Regression: Nonparametric Technology

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Export status	-0.0012	0.025*	0.027***	0.054***				
	(-0.15)	(2.46)	(3.38)	(4.92)				
Both					0.001	0.030*	0.036***	0.061***
					(0.04)	(2.48)	(3.59)	(4.74)
Exporter only					0.005	0.028	0.018	0.048*
					(0.49)	(1.53)	(1.48)	(2.50)
Importer only					0.012	0.018*	0.016*	0.015
					(1.86)	(2.05)	(2.14)	(1.45)
Fixed effects	Yes	Yes	No	No	Yes	Yes	No	No
R&D firms	Yes	No	Yes	No	Yes	No	Yes	No
Observations	24,222	12,170	24,222	12,170	24,222	12,170	24,222	12,170
R^2	0.885	0.889	0.866	0.871	0.885	0.889	0.866	0.871

t statistics in parentheses. Standard errors are clustered at the firm-level.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 5 presents the results for the linear cost share regression with a third order polynomial in the factor inputs. In the absence of fixed effects, there are large, positive, and statistically significant markup premia associated with exporting. For example, in the sample excluding R&D firms, exporters charge markups that are, on average, 5.4% higher than non-exporters. Including fixed effects reduces the size of these premia to 2.5%. There are also positive markup premia associated with each of the three trade classifications. The largest markup premia are among the set of firms that are both exporters and importers.

6 Conclusion

This paper makes two contributions. First, we perform a detailed assessment of an identification problem with the production approach to markups of De Loecker and Warzynski (2012) and De Loecker, Eeckhout, and Unger (2019). This problem arises when the empirical measure of output in the production function is deflated revenue. We extend the work of Klette and Griliches (1996), who concentrated on a Cobb-Douglas production technology, a CES demand system, and monopolistic competition, to assess the implications for the production approach of the omitted price bias in a more general setting. We show that the ratio of the revenue output elasticity and the cost share equals one for all firms and time periods. The assumption of monopolistic competition is essential for this result.

Second, we propose an econometric methodology that identifies firm heterogeneity in markups and averts the omitted price bias by not relying on production function estimation. This is our novel cost share approach that does not impose any structural assumptions beyond those of the production approach. The caveat of the cost share approach is that it does not identify the level of the markup. We apply our cost share approach to estimate the markup premia of exporters and importers using UK micro data. We find that there are positive, statistically significant, and economically meaningful markup premia for exporters. The evidence on the markup premia of importers is mixed. In subsequent work, we aim to explore the implications of the omitted price bias under other models of competition, such as Bertrand-Nash oligopoly.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). "Identification Properties of Recent Production Function Estimators". In: *Econometrica* 83, pp. 2411–2451.

Aghion, P. et al. (2018). "The Impact of Exports on Innovation: Theory and Evidence". In: *NBER Working Paper No. 24600.*

Arellano, M. and S.R. Bond (1991). "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations". In: *Review of Economic Studies* 58.2, pp. 277–297.

Basu, S. (2019). "Are Price-Cost Markups Rising in the United States? A Discussion of the Evidence". In: *Journal of Economic Perspectives* 33.3, pp. 3–22.

Bernard, A.B. and J.B. Jensen (1995). "Exporters, Jobs, and Wages in U.S. Manufacturing: 1976-87". In: *Brookings Papers on Economic Activity: Microeconomics.*

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

De Loecker, J. (2011). "Product Differentiation, Multiproduct Firms, and Estimating the Impact of Trade Liberalization on Productivity". In: *Econometrica* 79.5, pp. 1407–1451.

De Loecker, J., J. Eeckhout, and G. Unger (2019). "The Rise of Market Power and the Macroeconomic Implications". In: *Quarterly Journal of Economics (forthcoming).*

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

Foster, L., J. Haltiwanger, and C. Syverson (2008). "Reallocation, Firm Turnover, and Efficiency: Selection on Productivity or Profitability?" In: *American Economic Review* 98.1, pp. 394–425.

Hall, R. E. (1988). "The Relation Between Price and Marginal Cost in US Industry". In: *Journal of Political Economy* 96, pp. 921–947.

Hashemi, A. (2019). "Markups and Multidimensional Firm Heterogeneity". In: *M.Phil. thesis, University of Oxford.*

Klette, T.J. and Z. Griliches (1996). "The Inconsistency of Common Scale Estimators When Output Prices Are Unobserved and Endogenous". In: *Journal of Applied Econometrics* 11.4, pp. 343–361.

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

Marschak, J. and W. Andrews (1944). "Random Simultaneous Equations and the Theory of Production". In: *Econometrica* 12, pp. 143–205.

Melitz, M. (2003). “The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity”. In: *Econometrica* 71.6, pp. 1695–1725.

Melitz, M. and G.I.P. Ottaviano (2008). “Market Size, Trade, and Productivity”. In: *Review of Economic Studies* 75, pp. 295–316.

Olley, G. S. and A. Pakes (1996). “The Dynamics of Productivity in the Telecommunications Equipment Industry”. In: *Econometrica* 64, pp. 1263–1297.

ONS (2017). “Annual Respondents Database X, 1998-2014.” In: *Secure Access. [data collection]. 4th Edition. Office for National Statistics. SN: 7989.*

— (2019). “Business Expenditure on Research and Development, 1995-2017.” In: *Secure Access. [data collection]. 8th Edition. UK Data Service. SN: 6690.*

Robinson, P.M. (1988). “Root-N-Consistent Semiparametric Regression”. In: *Econometrica* 56, pp. 931–954.

Syverson, C. (2019). “Macroeconomics and Market Power: Context, Implications, and Open Questions”. In: *Journal of Economic Perspectives.*

Trushin, E. et al. (2015). “Research and Development Expenditures and Subsidies, Firm Productivity, Employment Creation and Survival, 1997-2012.” In: *Secure Access. [data collection]. UK Data Service. SN: 7716.*

A Appendix

A.1 Proof of Lemma 1

Proof. In each period t, the monopolist chooses its output Q_{it} to maximize its profit subject to its demand constraint. Denote by $P_{it}(Q_{it})$ the inverse demand function associated with (10) and by c_{it} the firm's marginal cost. The profit maximizing condition equates marginal revenue to marginal cost

$$P_{it}(Q_{it}) + Q_{it}\frac{\partial P_{it}(Q_{it})}{\partial Q_{it}} = c_{it}$$

Let

$$\epsilon_{it}^{d} \equiv -\frac{P_{it}(Q_{it})}{Q_{it}}\frac{\partial Q_{it}}{\partial P_{it}(Q_{it})}$$

denote the price elasticity of demand. The monopolist's optimality condition can be written in terms of the markup μ_{it} and the demand elasticity ϵ_{it}^{d}

$$\mu_{it} \equiv \frac{P_{it}(Q_{it})}{c_{it}} = \frac{\epsilon_{it}^{d}}{\epsilon_{it}^{d} - 1}$$

Under a CES demand system, the price elasticity of demand is constant and given by $\epsilon_{it}^{d} = \sigma$. This yields the desired result

$$\mu_{it} = \frac{\sigma}{\sigma - 1}$$

□

A.2 Proof of Proposition 2

Proof. The production technology is

$$Q_{it} = Q(\Omega_{it}, \boldsymbol{V_{it}}, K_{it}; \boldsymbol{\theta})$$

where $\boldsymbol{V_{it}} = \left(V_{it}^{1}, ..., V_{it}^{J}\right)$ are the J perfectly flexible inputs, K_{it} is the capital stock, Ω_{it} is unobserved firm productivity, and $\boldsymbol{\theta}$ is a vector of technology parameters. The single product monopolist faces a downward sloping inverse demand function $P_{it}(Q_{it})$. By definition of total revenue,

$$R_{it} = P_{it}(Q_{it})\, Q_{it}$$

Consider an arbitrary flexible input v. Then,

$$\frac{\partial R_{it}}{\partial V_{it}^{v}} = \frac{\partial Q_{it}}{\partial V_{it}^{v}} \left[P_{it}\left(Q_{it}\right) + Q_{it}\frac{dP_{it}\left(Q_{it}\right)}{dQ_{it}} \right] \tag{16}$$

Multiply both sides of equation (16) by V_{it}^{v}/R_{it} to give the revenue output elasticity

$$\gamma_{it}^{v} \equiv \frac{V_{it}^{v}}{R_{it}}\frac{\partial R_{it}}{\partial V_{it}^{v}} = \frac{V_{it}^{v}}{Q_{it}}\frac{\partial Q_{it}}{\partial V_{it}^{v}} \left[1 + \frac{Q_{it}}{P_{it}}\frac{dP_{it}\left(Q_{it}\right)}{dQ_{it}} \right]$$

The physical output elasticity is

$$\theta_{it}^{v} \equiv \frac{V_{it}^{v}}{Q_{it}}\frac{\partial Q_{it}}{\partial V_{it}^{v}}$$

and the price elasticity of demand is

$$\epsilon_{it}^{d} \equiv -\frac{P_{it}}{Q_{it}}\frac{dQ_{it}}{dP_{it}\left(Q_{it}\right)}$$

Then, we may express γ_{it}^{v} as a function of θ_{it}^{v} and ϵ_{it}^{d}

$$\gamma_{it}^{v} = \theta_{it}^{v}\left(\frac{\epsilon_{it}^{d} - 1}{\epsilon_{it}^{d}}\right) \tag{17}$$

Under monopolistic competition, the markup is identified as a function of the demand elasticity

$$\mu_{it} = \frac{\epsilon_{it}^{d}}{\epsilon_{it}^{d} - 1} \tag{18}$$

Combining equations (17) and (18) yields the following relationship between the revenue elasticity, the physical output elasticity, and the markup

$$\gamma_{it}^{v} = \frac{\theta_{it}^{v}}{\mu_{it}}$$

Lastly, the first order condition in cost minimization gives

$$\mu_{it} = \frac{\theta_{it}^{v}}{\alpha_{it}^{v}}$$

Suppose that the revenue elasticity were use in place of the physical output elasticity in equation (2). Then, the ratio of interest is

$$\begin{aligned}\frac{\gamma_{it}^x}{\alpha_{it}^x} &= \frac{\theta_{it}^x}{\alpha_{it}^x}\frac{1}{\mu_{it}} \\ &= \mu_{it}\frac{1}{\mu_{it}} \\ &= 1\end{aligned}$$

which yields the desired result.

$\square$