Online Appendices

Appendix A. Appendix for Section 2

Appendix A.1. Oligopolistic competition

The demand system, expressed in logarithms, facing an oligopolist i in period t is of the form

$$q_{it} = \mathcal{Q}_i\left(p_{it}, z_t\right) \tag{A.1}$$

where q_{it} and p_{it} denote the log output quantity and the log output price of firm i in period t, respectively, and z_t denotes the log industry expenditure function in period t that depends on the prices $(p_{1t}, \ldots, p_{Jt})$ of all N active firms operating in the industry in period t. Oligopolistic firms internalize the fact that they are nonatomistic in their industry and can influence industry expenditure z_t through their decisions. The key assumption underlying the demand system (A.1) is that the industry expenditure function z_t serves as a sufficient statistic for the prices of all competitors $\boldsymbol{p}_{-it} \coloneqq \{p_{kt}\}_{k \neq i}$ of firm i, given firm i's own price p_{it}.

The demand system (A.1) includes the popularized log-linear nested CES demand system of Atkeson and Burnstein (2008)

$$\mathcal{Q}_i\left(p_{it}, z_t\right) = \left(\rho - \eta\right) z_t - \rho p_{it}$$

where $\eta \geq 1$ denotes the constant elasticity of substitution of goods across industries and $\rho > \eta$ denotes the constant elasticity of substitution of goods within an industry. We do not restrict our subsequent analysis to this particular parametric functional form and instead work with the general demand system in equation (A.1).

Shephard's lemma (an envelope condition) identifies the partial change in z_t with respect to p_{it} as

$$\frac{\partial z_t}{\partial p_{it}} = S_{it}$$

where firm i's market share of the total industry revenues is

$$S_{it} \coloneqq \frac{P_{it} Q_{it}}{\sum_{k=1}^{J} P_{kt} Q_{kt}}$$

The own-price η_{it} and cross-price $\delta_{ik,t}$ elasticities of demand are defined as

$$\begin{aligned}
\eta_{it} \coloneqq -\frac{dq_{it}}{dp_{it}} &= -\left[\frac{\partial q_{it}}{\partial p_{it}} + \frac{\partial q_{it}}{\partial z_t} S_{it}\right], \\
\delta_{ik,t} \coloneqq \frac{dq_{it}}{dp_{kt}} &= \frac{\partial q_{it}}{\partial z_t} S_{kt}, \forall k \neq i
\end{aligned}$$

Totally differentiating the demand system (A.1) and the accounting identity for revenue $r_{it} = p_{it} + q_{it}$, and invoking Shephard's lemma, yields an expression for the elasticity of revenue r_{it} with respect to output q_{it}

$$\frac{dr_{it}}{dq_{it}} = \left(1 + \left[\frac{\partial q_{it}}{\partial p_{it}} + \frac{\partial q_{it}}{\partial z_t}\left(S_{it} + \sum_{k \neq i} S_{kt}\frac{dp_{kt}}{dp_{it}}\right)\right]^{-1}\right)$$

We emphasize that the elasticity dr_{it}/dq_{it} is different under the Bertrand and Cournot models of oligopolistic competition. The reason is that the firm's conjectural elasticities, $dp_{kt}/dp_{it}, \forall k \neq i$, differ between these two models of oligopoly. A useful benchmark is monopolistic competition, under which there are no strategic considerations, and therefore we obtain the simplification $dp_{kt}/dp_{it} = 0, \forall k \neq i$. This yields the familiar result $dr_{it}/dq_{it} = \frac{\eta_{it}-1}{\eta_{it}}$.

Recall that the revenue elasticity of the flexible input X^j_{it} is

$$\theta^{R,j}_{it} = \frac{dr_{it}}{dq_{it}}\theta^{Q,j}_{it}$$

Then, the estimand of the ratio estimator using the revenue elasticity in the numerator is

$$\begin{aligned}
\mu^R_{it} &= \frac{\theta^{R,j}_{it}}{\alpha^j_{it}} \\
&= \frac{dr_{it}}{dq_{it}}\frac{\theta^{Q,j}_{it}}{\alpha^j_{it}} \\
&= \frac{dr_{it}}{dq_{it}}\mu_{it}
\end{aligned}$$

Amiti et al. (2019) show that the firm's first order condition in the static profit maximization problem uniquely characterizes the firm's markup μ_{it} as a function of the firm's perceived price elasticity of demand σ_{it}. That is,

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

The perceived demand elasticity σ_{it} differs under Bertrand and Cournot competition. We now consider each in turn.

The Bertrand-Nash equilibrium condition is that all competitors of firm i hold their prices fixed, i.e. $dp_k = 0, \forall k \neq i$. Then, the elasticity of revenue with respect to output simplifies to

$$\frac{dr_{it}}{dq_{it}} = \frac{\eta_{it} - 1}{\eta_{it}}$$

The perceived demand elasticity under Bertrand competition is equal to the own-price demand elasticity.

$$\sigma_{it} = \eta_{it}$$

Then, the Bertrand markup $\mu_{it}^{Bertrand}$ is

$$\begin{aligned}\mu_{it}^{Bertrand} &= \frac{\eta_{it}}{\eta_{it}-1} \\ &= \left[\frac{dr_{it}}{dq_{it}}\right]^{-1}\end{aligned}$$

It follows that the estimand of the ratio estimator using the revenue elasticity does not identify the markup:

$$\begin{aligned}\mu_{it}^{R,Bertrand} &= \frac{dr_{it}}{dq_{it}}\mu_{it}^{Bertrand} \\ &= \frac{dr_{it}}{dq_{it}}\left[\frac{dr_{it}}{dq_{it}}\right]^{-1} \\ &= 1\end{aligned}$$

The Cournot-Nash equilibrium condition is that competitors hold their quantities fixed, i.e. $dq_k = 0, \forall k \neq i$. Then, the elasticity of revenue with respect to output simplifies to

$$\frac{dr_{it}}{dq_{it}} = \left(1 + \left[\frac{\partial q_{it}}{\partial p_{it}} + \frac{\partial q_{it}}{\partial z_t}\left(\frac{S_{it}}{1-\tilde{S}_{-i}}\right)\right]^{-1}\right)$$

where the response of competitiors is summarized in the statistic

$$\tilde{S}_{-i} := -\sum_{k\neq i}\left(\frac{\partial q_{kt}}{\partial p_{kt}}\right)^{-1}\left(\frac{\partial q_{kt}}{\partial z_t}\right)S_{kt}$$

The perceived demand elasticity under Cournot competition is equal to

$$\sigma_{it} = -\left[\frac{\partial q_{it}}{\partial p_{it}} + \frac{\partial q_{it}}{\partial z_t}\left(\frac{S_{it}}{1-\tilde{S}_{-i}}\right)\right]$$

Then, the Cournot markup $\mu_{it}^{Cournot}$ is

$$\begin{aligned}\mu_{it}^{Cournot} &= \frac{\sigma_{it}}{\sigma_{it}-1} \\ &= \left(1 + \left[\frac{\partial q_{it}}{\partial p_{it}} + \frac{\partial q_{it}}{\partial z_t}\left(\frac{S_{it}}{1-\tilde{S}_{-i}}\right)\right]^{-1}\right)^{-1} \\ &= \left(\frac{dr_{it}}{dq_{it}}\right)^{-1}\end{aligned}$$

Combining everything together, we again establish that the ratio estimator using the revenue elasticity does not identify the markup:

$$\begin{aligned}\mu_{it}^{R,Cournot} &= \frac{dr_{it}}{dq_{it}}\mu_{it}^{Cournot} \\ &= \frac{dr_{it}}{dq_{it}}\left(\frac{dr_{it}}{dq_{it}}\right)^{-1} \\ &= 1\end{aligned}$$

Appendix A.2. Input adjustment costs

We consider the same firm problem from Section 2, but we now assume that each input j is associated with a baseline quantity $\overline{X}_{it}^{j}$ and that the firm incurs adjustment costs when it chooses an input quantity $X_{it}^{j} \neq \overline{X}_{it}^{j}$. The baseline quantity $\overline{X}_{it}^{j}$ might reflect the input choice from the previous period in a dynamic version of the model. For simplicity, we assume that these costs are given by the smooth convex function $\kappa^{j}\left(X_{it}^{j}\right)$, which satisfies $\kappa^{j}\left(\overline{X}_{it}^{j}\right) = \frac{d\kappa^{j}\left(\overline{X}_{it}^{j}\right)}{dX_{it}^{j}} = 0$.

The firm's cost function is now given by

$$\begin{aligned}\mathcal{C}\left(Q_{it};\boldsymbol{W}_{t}\right) &:= \min_{\left\{X_{it}^{j}\right\}_{j=1}^{J}} \left\{\sum_{j=1}^{J} W_{t}^{j} X_{it}^{j} + \sum_{j=1}^{J} \kappa\left(X_{it}^{j}\right) W_{t}^{j}\right\} \\ \text{s.t.} &\quad \mathcal{F}\left(X_{it}^{1},\ldots,X_{it}^{J}\right) \geq Q_{it},\end{aligned}$$

where we have normalized the adjustment cost functions by the input price W_{t}^{j}. Following the same steps as in the previous section, we obtain the FOC

$$\frac{W_{t}^{j} X_{it}^{j}}{P_{it} Q_{it}}\left[1 + \frac{d\kappa^{j}\left(X_{it}^{j}\right)}{dX_{it}^{j}}\right] = \frac{\lambda_{it}}{P_{it}}\theta_{it}^{Q,j}.$$

Using α_{it}^{j} to denote the share of input j's cost in revenue and using the envelope condition, this implies

$$\alpha_{it}^{j}\left[1 + \frac{d\kappa^{j}\left(X_{it}^{j}\right)}{dX_{it}^{j}}\right] = \frac{\partial \mathcal{C}\left(\cdot\right)}{\partial Q_{it}}\frac{\theta_{it}^{Q,j}}{P_{it}}. \tag{A.2}$$

Hence, the ratio estimator using the revenue elasticity recovers

$$\mu_{it}^{R,j} = \frac{\theta_{it}^{R,j}}{\alpha_{it}^{j}} = 1 + \frac{d\kappa^{j}\left(X_{it}^{j}\right)}{dX_{it}^{j}},$$

and the ratio estimator using the output elasticity recovers

$$\mu_{it}^{Q,j}=\frac{\theta_{it}^{Q,j}}{\alpha_{it}^{j}}=\mu_{it}\left[1+\frac{d\kappa^{j}\left(X_{it}^{j}\right)}{dX_{it}^{j}}\right].$$

Why might it be more common to estimate $\mu_{it}^{R,j}>1$ than $\mu_{it}^{R,j}<1$ when using firm-level data? One hypothesis is that adjustment costs are asymmetrical. It is less costly to use less of an input than previously planned than to use more of an input. If this is the case then on average we would recover $\mu_{it}^{R,j}>1$. Similarly if firms are growing on average we would recover $\mu_{it}^{R,j}>1$ on average.

The argument above effectively assumes that observed input costs are $W_t^j X_{it}^j$ rather than $W_t^j X_{it}^j+W_t^j\kappa^j\left(X_{it}^j\right)$. If this is the measure of observed input costs then

$$\alpha_{it}^{j}=\frac{W_i^j X_{it}^j+W_t^j\kappa^j\left(X_{it}^j\right)}{P_{it}Q_{it}}$$

and we obtain

$$\frac{W_t^j X_{it}^j+W_t^j\frac{d\kappa^j\left(X_{it}^j\right)}{dX_{it}^j}}{P_{it}Q_{it}}=\frac{\lambda_{it}}{P_{it}}\theta_{it}^{Q,j}$$

$$\mu_{it}^{Q,j}=\frac{\theta_{it}^{Q,j}}{\alpha_{it}^{j}}=\mu_{it}\left(\frac{X_{it}^j+\frac{d\kappa^j\left(X_{it}^j\right)}{dX_{it}^j}}{X_{it}^j+\kappa^j\left(X_{it}^j\right)}\right)$$

so wedge >1 whenever $\kappa'>\kappa$.

Neither of the two cases that are typically considered in the literature lead to a bias. The flexible input case is $\kappa^j=0$, in which case the bias disappears. The fixed input case is one in which $X_{it}^j\to\overline{X}_{it}^j$ in which case the bias also disappears. (Note, however that the fixed input case is not the limit as $\kappa^j\to\infty$, and so is not a special case of the model with adjustment cost model. When $\kappa^j\to\infty$ in the adjustment cost model, the bias remains even in the limit, even though $X_{it}^j\to\overline{X}_{it}^j$).

Appendix A.3. Inputs that influence demand

In this section we show that even if output elasticities are available, markup estimates are biased whenever the variable factor of production is used partly to affect demand in addition to producing output.

We assume that the firm's production function is as in Section 2, but that its revenue is now given by

$$R_{it}\coloneqq\mathcal{P}\left(Q_{it},D_{it}\right)Q_{it}$$

where D_{it} is an endogenous demand shifter that the firm can influence through the use of inputs according to the function

$$D_{it} = \mathcal{D}\left(X_{it}^{D,1}, , \ldots, X_{it}^{D,J}\right).$$

We denote the amount of input j used in production as $X_{it}^{Q,j}$ and the amount used in influencing demand as $X_{it}^{D,j}$. The total quantity of input j used by the firm is $X_{it}^{j} = X_{it}^{Q,j} + X_{it}^{D,j}$.

The profit maximization problem of the firm is now

$$\Pi_{it} \quad := \quad \max_{Q_{it}, D_{it}} \left\{ \mathcal{P}\left(Q_{it}, D_{it}\right) Q_{it} - \mathcal{C}_Q\left(Q_{it}; \boldsymbol{W}_t\right) - \mathcal{C}_D\left(D_{it}; \boldsymbol{W}_t\right) \right\} \tag{A.3}$$

where $\mathcal{C}_Q\left(Q_{it}; \boldsymbol{W}_t\right)$ is the firm's cost function for producing output, defined by

$$\begin{aligned} \mathcal{C}_Q\left(Q_{it}; \boldsymbol{W}_t\right) \quad &:= \quad \min_{\left\{X_{it}^{Q,j}\right\}_{j=1}^{J}} \left\{ \sum_{j=1}^{J} W_t^j X_{it}^{Q,j} \right\} \\ &\text{s.t.} \quad Q_{it} \leq \mathcal{F}\left(X_{it}^{Q,1}, \ldots, X_{it}^{Q,J}\right) \end{aligned} \tag{A.4}$$

and $\mathcal{C}_D\left(D_{it}; \boldsymbol{W}_t\right)$ is the firm's cost function for influencing demand, defined by

$$\begin{aligned} \mathcal{C}_D\left(D_{it}; \boldsymbol{W}_t\right) \quad &:= \quad \min_{\left\{X_{it}^{D,j}\right\}_{j=1}^{J}} \left\{ \sum_{j=1}^{J} W_t^j X_{it}^{D,j} \right\} \\ &\text{s.t.} \quad \mathcal{D}\left(X_{it}^{D,1}, , \ldots, X_{it}^{D,J}\right) \geq D_{it} \end{aligned} \tag{A.5}$$

The optimality conditions from the profit maximization problem (A.3) are

$$1 - \frac{1}{\eta_{it}} = \frac{\partial \mathcal{C}_Q\left(\cdot\right)}{\partial Q_{it}} \frac{1}{P_{it}} \tag{A.6}$$

$$\varsigma_{it} = \frac{\partial \mathcal{C}_D\left(\cdot\right)}{\partial D_{it}} \frac{D_{it}}{P_{it} Q_{it}} \tag{A.7}$$

where ς_{it} describes the effect of the demand shifter on the price that a firm can charge for a given quantity of output. As in the previous section, the optimal markup of price over marginal production cost is

$$\mu_{it} := \left[\frac{\partial \mathcal{C}_Q\left(\cdot\right)}{\partial Q_{it}} \frac{1}{P_{it}}\right]^{-1} = \left(1 - \frac{1}{\eta_{it}}\right)^{-1}.$$

The FOC for the production cost minimization problem (A.4) yields the relationship

$$\alpha_{it}^{Q,j} = \frac{\partial \mathcal{C}_Q\left(\cdot\right)}{\partial Q_{it}} \frac{1}{P_{it}} \theta_{it}^{Q,j} \tag{A.8}$$

where $\alpha_{it}^{Q,j}$ is the share of revenue paid to input m for use in producing output, and $\theta_{it}^{Q,j}$ is the elasticity of output to the use of input j for production. It follows from equation (A.8) that if one could observe $X_{it}^{Q,j}$ separately from X_{it}^{j} then the ratio estimator would correctly recover the markup.

However, in practice we observe only the total usage of an input $X_{it}^{j} = X_{it}^{Q,j} + X_{it}^{D,j}$, rather then the usage in different activities separately. Using the FOC for the cost minimization problem for influencing demand (A.5) yields the relationship

$$\alpha_{it}^{D,j} = \frac{\partial \mathcal{C}_D\left(\cdot\right)}{\partial D_{it}} \frac{D_{it}}{P_{it}Q_{it}} \theta_{it}^{D,j} \tag{A.9}$$

where $\alpha_{it}^{D,j}$ is the share of revenue paid to input j for shifting demand and $\theta_{it}^{D,j}$ is the elasticity of D_{it} with respect to $X_{it}^{D,j}$. Combining (A.6),(A.7), (A.8) and (A.9) yields an expression for the total revenue share of input X_{it}^{j}

$$\alpha_{it}^{j} = \left(1 - \frac{1}{\eta_{it}}\right) \theta_{it}^{Q,j} + \varsigma_{it} \theta_{it}^{D,j} \tag{A.10}$$

To see what the ratio estimator recovers, note that the optimality condition for allocating an input j between producing goods $X_{it}^{Q,j}$ and influencing demand $X_{it}^{D,j}$ implies that the output elasticity of an input X_{it}^{j} is

$$\theta_{it}^{Q,j} \rho_{it}^{Q,j} + \frac{\partial \mathcal{F}}{\partial X_{it}^{D,j}} \frac{X_{it}^{D,j}}{Q_{it}} \psi_{it}^{D,j} = \theta_{it}^{Q,j} \rho_{it}^{Q,j} \tag{A.11}$$

where $\psi_{it}^{Q,j}$ is the elasticity of $X_{it}^{Q,j}$ with respect to X_{it}^{j} evaluated at the optimum. $\psi_{it}^{D,j}$ denotes the elasticity of $X_{it}^{D,j}$ with respect to X_{it}^{j} evaluated at the optimum. This means that in order to correctly recover the output elasticity of an input X_{it}^{j}, it is necessary to separately observe the part of that input that is actually used in producing goods as long as $\psi_{it}^{Q,j} \neq 1$. The fact that a firm uses inputs partly to influence demand introduces a bias into the estimate of the output elasticity. It also introduces a bias into the estimate of the markup. Combining (A.10) and (A.11) reveals that the estimand is given by

$$\mu_{it}^{Q,j} = \mu_{it} \left[\frac{\psi_{it}^{Q,j}}{1 + \frac{X_{it}^{D,j}}{X_{it}^{Q,j}}} \right].$$

There are however special cases in which $\psi_{it}^{Q,j} = 1$, i.e. the share of X_{it}^{j} in production and in influencing demand does not depend on the level of X_{it}^{j}. For example it is sufficient that the firm faces an isoelastic demand curve and $\mathcal{F}$ and $\mathcal{D}$ are Cobb-Douglas. If this is the case, there is no bias the estimate of the output elasticity, but the ratio estimator is still biased. [1]

$$\mu_{it}^{Q,j} = \mu_{it}\left[\frac{1}{1+\frac{X_{it}^{D,j}}{X_{it}^{Q,j}}}\right].$$

So if the flexible input is only used for production and not to influence demand ($X_{it}^{D,j} = 0$) then the ratio estimator recovers the markup. But if some of the input is used to influence demand, and this component is not separated out, then the ratio estimand will be biased. If, over time, the input X_{it}^{j} is increasingly being used to influence demand, then the ratio estimand will fall over time, without any change in the true markup.

Casual observation suggests that at least some part of the workforce currently employed in the corporate sector devotes its energy to influencing demand rather than to producing goods. This suggests that using labor as an input for estimating markups will yield estimates that are hard to interpret. When using the ratio estimator, heterogeneity across firms and industries in the extent to which they use labor for production versus marketing and sales-related expenses will thus manifest as heterogeneity in measured markups.

These observations also help shed light on the difference in the trend in markups that one obtains from Compustat data on US firms when one uses only COGS versus when one includes SGA as the flexible input (De Loecker et al., 2020; Traina, 2018). It seems reasonable to assume that in the COGS bundle, a larger fraction of the inputs is used to produce output and a smaller fraction is used to influence demand, than in the SGA bundle. Thus the downward bias in the ratio estimand is likely to be larger when including SGA in the bundle of flexible inputs, versus when using only COGS. Since the cost share of SGA in total revenue has been increasing relative to the cost share of COGS in total revenue, this will manifest as a widening gap between the ratio estimator that uses only COGS and the ratio estimator that also includes SGA. This is precisely what the literature has found.

So far in this section we have proceeded as if output were observed. If only revenue were observed, as in Section 2.1, then the ratio estimator again recovers $\mu_{it}^{R,j} = 1$, regardless of whether the input is being used for production or to influence demand. Given that Compustat data contains only revenue, not output, the aforementioned discussion is relevant only if one believes that the procedures in those papers do successfully recover output elasticities, which we believe they do not.

[1] This result does not require that $X_{it}^{D,j}$ and $X_{it}^{Q,j}$ are perfect substitutes, but it does require that they satisfy $X_{it} = h\left(X_{it}^{Q,j}, X_{it}^{D,j}\right)$ where h is a constant-returns-to-scale function. Thanks to Agustin Gutierrez for pointing this out.

Appendix B. Appendix for Section 3

Appendix B.1. Optimal input demand functions

This appendix supplies the derivations of the optimal input demand equation for intermediate inputs under two production technology specifications. Section Appendix B.1.1 provides the derivation for a Cobb-Douglas technology, while Section Appendix B.1.2 provides that for a non-parametric technology.

Appendix B.1.1. Cobb-Douglas technology

The three-factor Cobb-Douglas production function for gross output Q_{it} with Hicks-neutral productivity ω_{it} is

$$Q_{it} = \exp\left(\omega_{it}\right) K_{it}^{\beta_K} L_{it}^{\beta_L} M_{it}^{\beta_M}$$

Since M_{it} is the single flexible input, the cost minimizing input demand for M_{it} can be obtained by rearranging the Cobb-Douglas production function conditional on a given output quantity Q_{it}

$$M_{it} = M^{*}\left(Q_{it}; K_{it}, L_{it}, \omega_{it}\right) \coloneqq \exp\left(-\frac{1}{\beta_M}\omega_{it}\right) Q_{it}^{\frac{1}{\beta_M}} K_{it}^{-\frac{\beta_K}{\beta_M}} L_{it}^{-\frac{\beta_L}{\beta_M}} \tag{B.1}$$

Then, the minimized total variable cost function is

$$\mathcal{C}\left(Q_{it}; K_{it}, L_{it}, P_{it}^{M}, \omega_{it}\right) \coloneqq P_{it}^{M} M^{*}\left(Q_{it}; K_{it}, L_{it}, \omega_{it}\right) \tag{B.2}$$

where P_{it}^{M} is the unit input price of M_{it} that firm i takes as given. Taking the demand system $P_{it} = \mathcal{P}\left(Q_{it}\right)$ and the total cost function $\mathcal{C}\left(Q_{it}; K_{it}, L_{it}, P_{it}^{M}, \omega_{it}\right)$ as given, firm i chooses Q_{it} to solve a static profit maximization problem

$$\max_{Q_{it}} \left\{\mathcal{P}\left(Q_{it}\right) Q_{it} - \mathcal{C}\left(Q_{it}; K_{it}, L_{it}, P_{it}^{M}, \omega_{it}\right)\right\}$$

The first order condition in profit maximization equates marginal revenue to marginal cost

$$\mathcal{P}\left(Q_{it}\right)\left(\frac{\eta_{it}-1}{\eta_{it}}\right) = \frac{\partial \mathcal{C}\left(Q_{it}; K_{it}, L_{it}, P_{it}^{M}, \omega_{it}\right)}{\partial Q_{it}} \tag{B.3}$$

where η_{it} is the absolute value of the price elasticity of demand. Equation (B.3) identifies the markup μ_{it} under monopolistic competition as a function of the demand elasticity.

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

Applying the functional form in equation (B.1) to the FOC in equation (B.3) and solving for $q_{it} \coloneqq \ln Q_{it}$ gives

$$q_{it} = \frac{\beta_M}{1-\beta_M}\ln\beta_M + \frac{\beta_k}{1-\beta_M}k_{it} + \frac{\beta_l}{1-\beta_M}l_{it} + \frac{\beta_M}{1-\beta_M}\left(p_{it} - \ln\mu_{it} - p_{it}^{M}\right) + \frac{1}{1-\beta_M}\omega_{it} \tag{B.4}$$

where $p_{it}^{M} \coloneqq \ln P_{it}^{M}$ and $p_{it} \coloneqq \ln P_{it}$. Using equation (B.4) to substitute for q_{it} in equation (B.1) produces the optimal input demand equation for m_{it} in terms of the state variables $(k_{it}, l_{it}, \omega_{it})$, the exogenous input price p_{it}^{M}, and the endogenous optimal output price p_{it} and markup μ_{it}.

$$m_{it} = \frac{\ln\beta_M}{1-\beta_M} + \frac{\beta_K}{1-\beta_M} k_{it} + \frac{\beta_L}{1-\beta_M} l_{it} + \frac{1}{1-\beta_M}\left(p_{it} - \ln\mu_{it} - p_{it}^{M} + \omega_{it}\right)$$

Appendix B.1.2. Non-parametric technology

The non-parametric three-factor production function for gross output with productivity ω_{it} is

$$Q_{it} = \mathcal{F}\left(K_{it}, L_{it}, M_{it}, \omega_{it}\right) \tag{B.5}$$

The only restriction we impose on the function $\mathcal{F}(\cdot)$ is that it is twice continuously differentiable in each of its arguments. As in Section Appendix B.1.1, M_{it} is the single flexible input. Inverting equation (B.5) produces the cost-minimizing input demand for M_{it}.

$$M_{it}^{*} = \mathcal{F}^{-1}\left(Q_{it}; K_{it}, L_{it}, \omega_{it}\right) \tag{B.6}$$

The minimized total variable cost function is thus

$$\mathcal{C}\left(Q_{it}; K_{it}, L_{it}, P_{it}^{M}, \omega_{it}\right) \coloneqq P_{it}^{M} \mathcal{F}^{-1}\left(Q_{it}; K_{it}, L_{it}, \omega_{it}\right)$$

The first order condition in profit maximization is then

$$\mathcal{P}\left(Q_{it}\right)\left(\frac{\eta_{it}-1}{\eta_{it}}\right) = P_{it}^{M} \frac{\partial \mathcal{F}^{-1}\left(Q_{it}; K_{it}, L_{it}, \omega_{it}\right)}{\partial Q_{it}} \tag{B.7}$$

Given a functional form for $\mathcal{F}(\cdot)$, one can solve equation (B.7) for the optimal output level Q_{it}.

$$Q_{it} = Q^{*}\left(K_{it}, L_{it}, P_{it}^{M}, \omega_{it}, P_{it}, \mu_{it}\right) \tag{B.8}$$

Using equation (B.8) to substitute for Q_{it}^{*} in equation (B.6) yields the optimal input demand function for intermediate inputs.

$$\begin{aligned} M_{it} &= \mathcal{F}^{-1}\left(Q^{*}\left(K_{it}, L_{it}, P_{it}^{M}, \omega_{it}, P_{it}, \mu_{it}\right); K_{it}, L_{it}, \omega_{it}\right) \\ &\coloneqq M^{*}\left(K_{it}, L_{it}, P_{it}^{M}, \omega_{it}, P_{it}, \mu_{it}\right) \end{aligned}$$

In the absence of price data on inputs and outputs, the scalar unobservables in the input demand function $M^{*}(\cdot)$ are $\left(P_{it}^{M}, \omega_{it}, P_{it}, \mu_{it}\right)$.

Appendix B.2. Learning about variation in markups from variation in the cost share only

Without a way to estimate the output elasticity for a flexible input consistently from typical production data, we cannot use the ratio estimator to learn about the level of markups. We can however still use insights from the production approach to learn about variation in markups across firms. This variation can be studied using a regression model for the log of the cost share in total revenue for a perfectly flexible input. We sketch this 'cost share approach' to studying markups in this appendix.

As discussed in Section 2, the ratio estimator relies on the relationship $\mu_{it} = \frac{\theta_{it}^{Q,j}}{\alpha_{it}^{j}}$ for a flexible input X_{it}^{j}. Taking logs and rearranging, we obviously have $-\ln \alpha_{it}^{j} = -\ln \theta_{it}^{Q,j} + \ln \mu_{it}$. First consider the three factor, Cobb-Douglas case in which intermediate inputs (M_{it}) is the perfectly flexible input, as discussed in Section 3. Here $\ln \alpha_{it}^{M} = (p_{it}^{M} + m_{it}) - (p_{it} + q_{it})$ is the log of the true cost share in revenue for intermediate inputs, and $\ln \theta_{it}^{M} = \ln \beta_{M}$ is a constant term. Letting $\ln s_{it}^{M} = (p_{it}^{M} + m_{it}) - (p_{it} + y_{it})$ denote the log of the observed cost share in revenue for firm i in period t, we then have

$$-\ln s_{it}^{M} = -\ln \beta_{M} + \ln \mu_{it} + \varepsilon_{it} \tag{B.9}$$

where $y_{it} = q_{it} + \varepsilon_{it}$ as before.[2]

Without a consistent estimate of the output elasticity (β_{M}), it is clear that the mean of the log of the observed cost shares conflates the log of the output elasticity and the mean of the log of the markups, and does not separately identify the latter. Nevertheless, under the maintained assumption that the output elasticity is common to all the firm-year observations, we can use this relation to study variation in markups. For example, if the binary dummy D_{it}^{X} indicates whether or not firm i in period t is an exporter, we can specify a linear relationship between log markups and export status

$$\ln \mu_{it} = \delta_{0} + \delta_{1} D_{it}^{X} + \nu_{it} \tag{B.10}$$

as in De Loecker and Warzynski (2012). Substituting (B.10) into (B.9), we have the linear specification

$$-\ln s_{it}^{M} = (\delta_{0} - \ln \beta_{M}) + \delta_{1} D_{it}^{X} + (\varepsilon_{it} + \nu_{it}) \tag{B.11}$$

In the Cobb-Douglas case, we can thus learn about the *association* between log markups and export status from a simple regression of the log of the observed cost share in revenue for a flexible input on a constant and the export status dummy.[3]

For more general Hicks-neutral gross output production functions, we can write the

[2]For simplicity, we assume here that this is the only source of measurement error in the log of the observed cost share in revenue. In the Cobb-Douglas case, we can easily allow for (multiplicative) measurement error in both the numerator and the denominator of the cost share for intermediate inputs.

[3]As in De Loecker and Warzynski (2012), additional controls can be included in this regression specification, but OLS is still unlikely to estimate the causal effect of exporting on markups consistently. If the sample used to estimate (B.11) pools data for firms in several sectors, sector dummies can be used to allow for heterogeneity in the output elasticity β_{M} between sectors.

log of the output elasticity $\ln \theta_{it}^M = f(k_{it}, l_{it}, m_{it})$,[4] in which case (B.11) becomes

$$-\ln s_{it}^M = g(k_{it}, l_{it}, m_{it}) + \delta_1 D_{it}^X + (\varepsilon_{it} + \nu_{it}) \tag{B.12}$$

where $g(k_{it}, l_{it}, m_{it}) = \delta_0 - f(k_{it}, l_{it}, m_{it})$. We can then learn about the association between log markups and export status either by approximating $g(k_{it}, l_{it}, m_{it})$ using a flexible functional form, or by estimating (B.12) using semi-parametric methods for partially linear models (Robinson, 1988).

This cost share approach allows us to learn about some forms of variation across firms in markups under essentially the same assumptions needed for the production approach, but without requiring a consistent estimate of the output elasticity. Except in the Cobb-Douglas case, we could not use this approach to study the association between markups and measures of firm size (e.g. the log of employment, l_{it}) or measures of factor intensity (e.g. the log of the capital-labor ratio, $k_{it} - l_{it}$); we may also have low power to detect significant association between markups and observed firm characteristics that are strongly correlated with functions of the production inputs. In principle, this approach could also be used to study trends in markups over time, as in De Loecker et al. (2020). However, it should be emphasized that the trend in the log of the cost share in revenue for a flexible input identifies the trend in the log of the markup only under the maintained assumption that the output elasticity is stable over time, which cannot be verified without a way of estimating the output elasticity consistently for different sub-periods.

[4]For example, in the translog case, we have $f(k_{it}, l_{it}, m_{it}) = \ln(\beta_M + \beta_{KM} k_{it} + \beta_{LM} l_{it} + \beta_{MM} m_{it})$.