Partial identification of production functions with flexible inputs

Zach Flynn and Amit Gandhi

December 18, 2018

Abstract

We estimate markups from production data by directly recovering the production function and inferring marginal costs. The challenge in doing so is that we must first identify the production function, itself a difficult identification problem. Proxy methods and other common structural methods of identifying the production function make assumptions that *restrict* markups to identify the production function so they can not be used to *infer* markups from the production function. We show how to modify the standard proxy model to avoid restricting markups while still being able to partially identify the production function and use this approach to put meaningful bounds on the relationship between export status and markups.

1 Introduction

Measuring markups is a classic problem in empirical industrial organization. The New Empirical Industrial Organization (NEIO) paradigm attacks the problem through the demand side, combining an estimated demand system with first order conditions from a model of competition to recover markups. But when data on demand is not rich, the approach can not be used as intended because it forces the researcher to use a restrictive demand system to augment the lack of variation in the data (say, the CES demand system). The functional form of demand, as opposed to variation in the data, then drives inferences about markups. The empirical strategy also requires the researcher to impose a model of firm conduct (which is often a variation of Bertrand-Nash pricing). Inferences about markups will be driven by this modeling choice when we have little prior reason to suppose any particular conduct model. The goal of learning the markup distribution is to learn something about market power within the industry which is strongly tied to the conduct of the industry. Can we learn about markups without assuming conduct?

We approach the problem, as De Loecker and Warzynski (2012) do, through the cost side of the firm's problem. If we can recover a firm's cost or production function without assuming conduct, then we can directly recover marginal cost without making assumptions about the form of the conduct part of the firm's problem (we can think of conduct as a function mapping firm choices to markups).

Cost minimization implies that for any "flexible" input — an input that has no dynamic consequences and can be freely adjusted — the following equation holds,

$$\frac{P}{\text{MC}} = \frac{\text{Revenue}}{\text{Expenditure on flexible input}} \times \text{Output elasticity of flexible input}, \tag{1}$$

where P is output price and MC is marginal cost. The equation suggests an empirical strategy. The left hand side of the equality is a measure of markups. If we can measure the output elasticity of a variable input from production data, without assuming conduct, then we know the right hand side of (1) and can recover markups. De Loecker and Warzynski (2012) suggest exactly this strategy.

The challenge is that identifying the output elasticity of a variable input from input and output data is itself nontrivial. Estimation of the production function faces a classic simultaneity problem — inputs are chosen on the basis of productivity, the "transmission bias" of Marschak and Andrews 1944 — and to overcome that problem, we often restrict markups.

De Loecker and Warzynski (2012) appeal (understandably) to the "proxy" literature for identification (see Olley and Pakes 1996, Levinsohn and Petrin 2003, Ackerberg, Caves, and Frazer 2015) which restricts the plant's input choice problem in order to allow observable variables to "proxy" for unobserved productivity.

But Gandhi, Navarro, and Rivers (2015) show the output elasticities of variable inputs are *not* identified under the economic restrictions imposed in the proxy literature[1]. We review the non-identification result and its implications for the markup problem in Section 2. In Section 8.3, we illustrate how we would make misleading inferences about markups if we used proxy variable estimators with a Monte Carlo.

Our constructive contribution is to show that equation (1) *can* be used to learn about markups despite the non-identification result in Gandhi, Navarro, and Rivers (2015). There are two key ideas: first, the proxy structure can be used to *partially* identify the production function without pinning down markups in spite of the non-*point* identification result in Gandhi, Navarro, and Rivers (2015); second, that it is not the markup of a particular firm that is of interest for policy purposes, but the *pattern* or distribution of markups in the industry. Ultimately, we are interested in some element of the vector of parameters β that characterize the linear relationship between markups on firm/plant characteristics (X),

$$\frac{P}{\text{MC}} = X^{\top}\beta + U, \quad \mathbb{E}[XU] = 0. \tag{2}$$

Although β is not point identified (because markups are not point identified), we show β can be partially identified using the economic restrictions of the proxy model. The upper and

[1] The empirical strategy proposed in Gandhi, Navarro, and Rivers (2015) to solve the nonidentification problem in fact goes in the opposite direction of De Loecker and Warzynski (2012). They show that if we know markups, then the flexible input elasticity can be identified using equation (1). With knowledge of the flexible input elasticity, the full production function can be recovered by using the standard structure in the proxy literature. But this precludes using equation (1) to estimate markups because we have already assumed markups to recover the production function.

lower bounds on β can be expressed as a linear program and there are useful, computationally-convenient methods to make inference on the value of those programs. We apply our identification strategy to production data from Chile and find that our bounds have enough content to sign the relationship between exporter status, markups, and productivity in many industries.

Standard trade models with firm heterogeneity—like Melitz and Ottaviano (2008)—predict that more productive firms will have higher markups within their own market. Because such models predict that exporters will be more productive, they also predict the domestic firms that engage in trade will have more market power. We find general support for the prediction across industries but only when we control for plant contact with the world market. Unconditional on plant contact with the world market, we would find a negative relationship because, we show, Chilean plants earn substantially lower markups in the world market than in their domestic market.

In Section 2, we describe the proxy model and methods of recovering markups based on it. We show that these methods can not identify markups without observing plant-level wages that can be excluded—differences in plant-level wages can not reflect differences in productivity.

In Section 3, we lay out our identification approach based on a weak version of the first order condition approach in Gandhi, Navarro, and Rivers (2015) and the standard proxy assumptions.

In Section 5, we show how to use our identification assumptions to form confidence intervals around parameters of interest.

In Section 9, we discuss the data we use to learn about the relationship between markups, plant size, and plant exporter status.

In Section 8.2, we present a simplified version of the model in Melitz and Ottaviano (2008) to show the economics behind the main theoretical prediction we test.

We define our measure of markups in Section 8.1, we present some "reduced-form" findings about the data in Section 10, and we give our main empirical estimates in Section 11.

2 The Model and Non-Identification Problem

Let q_t be log output and (ℓ_t, k_t, m_t) be the log of labor, capital, and materials respectively in period t. The data are inputs and output over the panel $t = 1, \ldots T$,

$$y = \left\{ \left(q_t, \ell_t, k_t, m_t \right) \right\}_{t=1}^{T}. \tag{3}$$

The joint distribution of the data y in the underlying population of plants is identified in the data. Capital letters stand for the levels of the logged variables so that, Q_t is output, L_t is labor, K_t is capital, and M_t is materials.

The proxy approach relates this data to a model of production with three main parts (we will call this set of assumptions, the "proxy structure"):

1. Output and inputs in each period are related in the following way:

$$q_t = f\left(\ell_t, k_t, m_t\right) + a_t + \epsilon_t, \tag{4}$$

where f is the production function characterizing the technology of an industry, a_t is a productivity shock that the firm observes before making its period t input decisions, and ϵ_t is an ex-post shock that is independent of the firm's input decisions.

2. Productivity is a Markov process,

$$a_t = g(a_{t-1}) + e_t, \tag{5}$$

where the shocks e_t are uncorrelated with inputs chosen before period t, so:

$$\mathbb{E}[e_t | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}, \dots] = 0, \tag{6}$$

where the ellipses represent all other lags of the inputs.

3. Plants choose materials m_t flexibly in each period and this choice does not have any dynamic implications. Capital and labor, (ℓ_t, k_t), are quasi-fixed: they are chosen before period t and treated as a state variable by the plant in period t (chosen before e_t is known by the plant).

   Material input demand has the form $m_t = m_t(\ell_t, k_t, a_t)$ where m_t is a strictly increasing function of a_t. The model is called the "proxy structure" because this assumption implies productivity can be proxied by observable inputs, $a_t = m_t^{-1}(m_t, \ell_t, k_t)$.

Gandhi, Navarro, and Rivers (2015) show this structure is insufficient to identify f. For any proposed production function f there exists an $\widetilde{f}$ that is observationally equivalent: both f and $\widetilde{f}$ are consistent with the proxy structure and generate the same joint distribution of observables. Any estimator motivated by restrictions generated by this structure will not be consistent (except for potentially arbitrary functional form restrictions on f). But Gandhi, Navarro, and Rivers (2015) show that if the flexible input elasticity,

$$\frac{\partial}{\partial m_t} f(\ell_t, k_t, m_t),$$

were identified through an outside source, then the production function f would be identified over the support of the data given the proxy structure; the source of under-identification is the flexible input elasticity.

Gandhi, Navarro, and Rivers (2015) solve the identification problem by assuming a markup ($P = \text{MC}$, say) and using the relationship between markups and the flexible output elasticity to recover the production function. But, if our purpose is to use the relationship (1) to learn about markups, we can not use this method.

De Loecker and Warzynski (2012) use lags of flexible inputs to instrument for current flexible inputs. Their argument is that lags of flexible inputs will be correlated with the price of the plant's inputs which will shift the distribution of current inputs but, since they were chosen in the past, they will be uncorrelated with later shocks to productivity. Their argument introduces input price variation — which is absent from the proxy structure.

If there is no wage variation, then Gandhi, Navarro, and Rivers (2015) have already shown the model is not identified. Briefly, without wage variation, the proxy model is,

$$q_{it} = f(\ell_{it}, k_{it}, m_{it}) + g(\ell_{it-1}, k_{it-1}, m_{it-1}) + e_{it} + \epsilon_{it} \tag{7}$$

$$\mathbb{E}[e_{it} + \epsilon_{it} | \ell_{it}, k_{it}, \ell_{it-1}, k_{it-1}, m_{it-1}, \ldots] = 0, \tag{8}$$

where the ellipses stands for all other lags of the inputs, i indexes plants, and t indexes the unit of time. The moment restriction is insufficient to identify the production function because, under the proxy model,

$$m_{it} = m_t(\ell_{it}, k_{it}, a_{it}) = m_t\left(\ell_{it}, k_{it}, g(\ell_{it-1}, k_{it-1}, m_{it-1}) + e_{it}\right), \tag{9}$$

so, conditional on $(\ell_{it}, k_{it}, \ell_{it-1}, k_{it-1}, m_{it-1})$, the only variation in m_{it} is via e_{it} but, by assumption, e_{it} can not be predicted by lagged inputs so lagged inputs are not valid instruments for m_{it}.

The difference between the model De Loecker and Warzynski (2012) have in mind and the proxy assumptions used in Gandhi, Navarro, and Rivers (2015) is that De Loecker and Warzynski (2012) allow input prices to vary by plant which gives another plant-specific state variable aside from e_{it} that might shift m_{it}. Does the presence of input price variation break the non-identification result?

There are two problems: the nature of the input price variation required for identification is more demanding than we might intuit (it is not enough that input prices vary by plant) and the nature of the data required is more demanding than we see in practice (we must observe the plant-level wages). We will argue these two problems demand a new approach that does not rely on this variation. But, before we present our arguments, we should acknowledge that using input prices for identification can work in some special situations, and it provides point identification while our methods will give only partial identification.

First, say input prices vary across plants, and we observe these input prices. In that case, material demand is,

$$m_t = m_t(a_t, \ell_t, k_t, w_t) = m_t\left(g(\ell_{t-1}, k_{t-1}, m_{t-1}, w_{t-1}) + e_t, \ell_t, k_t, w_t\right), \tag{10}$$

where w_t is the log price of materials.

The first question for identification is whether m_{t-2} has any strength as an instrument; is it correlated with m_t conditional on $(\ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}, w_{t-1})$? Aside from the variables that are conditioned on, there are two state variables that affect material demand: (e_t, w_t). By the proxy assumptions, m_{t-2} is not correlated with e_t so, for the instrument to have any strength, it must be correlated with w_t, *conditional on* $(\ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}, w_{t-1})$.

To see why this is a stronger assumption that we might think, suppose w_t is an AR1 process,

$$w_{it} = \rho^w w_{it-1} + \text{innovation}_{it}, \quad \text{innovation}_{it} \sim N(0,1). \tag{11}$$

where the innovation is a shock, independent of any decisions the plant makes. Then, conditional on w_{it-1}, the only variation in w_{it} is through the innovation term which is entirely independent of m_{it-2}. If this model is the data generating process, then m_{it-2} has no strength as an instrument. Similarly, if w_{it} were just fluctuations around a plant-specific mean — $w_{it} = \delta_i + \text{fluctuation}_{it}$, where the fluctuations are iid — the model would not be identified.

What we need for m_{it-2} to be an instrument is for there to be a plant-specific component of the wage process—for example,

$$w_{it} = \rho^w w_{it-1} + \xi_i + \text{innovation}_{it} \text{—} \tag{12}$$

or that w_t is at least an AR2 process; there needs to be some underlying state variable that affects w_{it} aside from w_{it-1} that m_{it-2} can proxy.

The second question for identification is whether m_{t-2} is a *valid* instrument. The question might seem to be beside the point because we have already assumed m_{t-2} is uncorrelated with the innovation term in the proxy structure we laid out above. But this structure is justified within a model where inputs are homogeneous and input price variation suggests input quality variation. If input price variation does reflect input quality variation, then innovations in the wage process are not unrelated to innovations in the productivity process—when wages go up, it is because the plant is using more productive inputs—in which case m_{t-2} is correlated with e_t.

But the larger issue is that we often do not observe w_t. If input prices vary across plants but we do not observe this variation, then using twice-lagged[2] flexible inputs as instruments will not work because then, the proxy function is,

$$m_t = m_t(a_t, \ell_t, k_t, w_t) \implies a_t = m_t^{-1}(m_t, \ell_t, k_t, w_t) \tag{13}$$

$$\implies q_t = f(\ell_t, k_t, m_t) + g(\ell_{t-1}, k_{t-1}, m_{t-1}, w_{t-1}) + e_t + \epsilon_t. \tag{14}$$

If we do not observe w_{t-1} and we run the above model, omitting w_{t-1}, we introduce an omitted variable bias. So the De Loecker and Warzynski (2012) approach requires that we observe wages, which we do not usually do.

In fact, the reason to use lagged flexible inputs as the instrument is that we do not observe plant wages—otherwise, if we believe wage variation is excluded, we could just use plant-level wages (w_{it}) directly as the instrument and it would likely be a stronger instrument than indirectly proxying w_{it} via m_{it-2}.

Only in the case where input prices vary in a particular way and where this variation is exogenous to productivity (some plants are stuck with getting a worse input price for similar inputs) can the method be used to recover the production function. Because we do not usually observe exogenous variation in plant-level input prices, an alternative method to recover markups would be useful.

[2] De Loecker and Warzynski (2012) use a value-added production function and treat labor as a flexible input. In that case, labor does not appear in the material demand function so first-lagged labor is excluded. But when materials are the flexible input and included in the production function, lagged materials are not excluded so we need twice-lagged flexible input use.

We develop one, showing how to use the standard proxy assumptions to partially identify the markup distribution and avoid these issues.

3 Partial identification of the production function in the proxy model

Our goal is to identify the flexible output elasticity without putting too much structure on conduct and to avoid the issues plaguing the use of input price variation for identification. Accomplishing this goal while maintaining point identification is difficult without making strong assumptions on exactly how plants choose inputs or on the functional form of the production function. We use a partial identification approach derived under the same assumptions as the proxy approach which allows for a broad class of conducts and technologies.

We show that while the assumptions of the proxy methods do not point identify the flexible output elasticity, they do partially identify it.

Identification of the markup distribution depends on identification of the material output elasticity. The proxy approach in Levinsohn and Petrin (2003) argues material demand is increasing in productivity. They use this assumption to argue productivity is a function of observed inputs, but the assumption also implies the inverse material demand function is increasing in materials, or:

$$a_t = m_t^{-1}(m_t, k_t, \ell_t), \quad \frac{\partial m_t^{-1}}{\partial m_t} \geq 0 \tag{15}$$

$$\implies \frac{\partial}{\partial m_t}\mathbb{E}\left(q_t | \ell_t, k_t, m_t\right) \geq \frac{\partial f}{\partial m}, \tag{16}$$

an upper bound on the flexible output elasticity. While this restriction is not usually imposed empirically in the proxy literature, it has identifying power and is not an extra assumption beyond the standard proxy structure introduced in Section 2.

For a lower bound on the flexible output elasticity, we weaken Gandhi, Navarro, and Rivers (2015)'s assumption that planned markups are zero, to the assumption that planned markups are greater than zero—this assumption is our one addition to the proxy structure, but it is satisfied in the standard models used to justify the assumption that material demand is increasing in productivity.

Planned output (Q^E) is the output the plant expects to produce, given its input choices and productivity,

$$Q^E = F(L, K, M)\, A \times \mathbb{E}\left[\exp(\epsilon)\right]. \tag{17}$$

While we do not observe planned output, we do observe output which is,

$$Q = F(L, K, M)\, A \exp(\epsilon) = Q^E \mathbb{E}\left[\exp(\epsilon)\right]^{-1} \exp(\epsilon) \tag{18}$$

$$\implies Q^E = Q \exp(-\epsilon)\, \mathbb{E}\left[\exp(\epsilon)\right]. \tag{19}$$

Because, by the proxy assumptions,

$$q_t = f(\ell_t, k_t, m_t) + a_t + \epsilon_t = f(\ell_t, k_t, m_t) + m_t^{-1}(m_t, \ell_t, k_t) + \epsilon_t \tag{20}$$

$$\mathbb{E}[\epsilon_t | \ell_t, k_t, m_t] = 0 \implies q_t = \mathbb{E}\left[q_t | \ell_t, k_t, m_t\right] + \epsilon_t, \tag{21}$$

we can recover ϵ_t from data and so, planned output (Q^E).

Our measure of planned markups is then, letting Revenue^E be expected revenue,

$$\frac{\text{Revenue}^E}{W_M M} \times \frac{\partial f}{\partial m} \geq 1 \iff \frac{\partial f}{\partial m} \geq \frac{W_M M}{\text{Revenue}^E} = \frac{W_M M}{\text{Revenue} \times \mathbb{E}\left[\exp(\epsilon)\right]} \times \exp(\epsilon) \tag{22}$$

$$\implies \frac{\partial f}{\partial m} \geq \exp\left\{\mathbb{E}\left[\log\left(\frac{W_M M}{\text{Revenue} \times \mathbb{E}\left[\exp(\epsilon)\right]}\right) | \ell, k, m\right]\right\}, \tag{23}$$

which gives a lower bound on the material output elasticity. The lower bound is exactly Gandhi, Navarro, and Rivers (2015)'s estimate of the material output elasticity.

With just these two assumptions, we can bound parameters of the markup distribution. These assumptions exhaust the power of the proxy structure to identify the markup distribution because these are the only restrictions the proxy model of production puts on the flexible output elasticity.

4 Revenue as output

We do not usually observe physical output. Instead, we observe revenue and we deflate revenue using various deflators. The mismeasurement of output can have serious consequences for production function estimates when there is imperfect competition — which is exactly what we are assuming when we are estimating markups.

While our approach is not entirely immune from this challenge, it is more resilient to using revenue as a measure of output than most common production function estimation approaches by virtue of the fact that the validity of the identified set does not depend on observing output.

Our lower bound on the output elasticity does not depend on output at all,

$$\frac{\partial f}{\partial m} \geq \exp\left\{\mathbb{E}\left[\log\left(\frac{W_M M}{\text{Revenue} \times \mathbb{E}\left[\exp(\epsilon)\right]}\right) | \ell, k, m\right]\right\}, \tag{24}$$

so we only need to observe revenue.

Our upper bound on the output elasticity is,

$$\frac{\partial}{\partial m}\mathbb{E}\left[q | \ell, k, m\right] \geq \frac{\partial f}{\partial m}, \tag{25}$$

so the upper bound does depend on physical output. Consider replacing physical output with revenue above so that,

$$\frac{\partial}{\partial m}\mathbb{E}\left[p + q | \ell, k, m\right] = \frac{\partial}{\partial m}\mathbb{E}\left[p | \ell, k, m\right] + \frac{\partial}{\partial m}\mathbb{E}\left[q | \ell, k, m\right] = \frac{\partial}{\partial m}\mathbb{E}\left[p + a | \ell, k, m\right] + \frac{\partial f}{\partial m} \tag{26}$$

If,

$$\frac{\partial}{\partial m}\mathbb{E}\left[p+a|\ell,k,m\right]\geq 0, \tag{27}$$

then we can use revenue in the place of output as the upper bound.

There is good reason to suppose that $\mathbb{E}\left[p+a|\ell,k,m\right]$ is increasing in m. Suppose that each plant faces a demand curve, $p_i=\xi_i+\eta q_i$ (where p is log output price), where the demand curve is elastic so that $0>\eta>-1$. The constant elasticity form is not crucial, but it helps to illustrate the main idea.

The static part of the profit maximization problem is,

$$\max_{M}\Xi_i F(L,K,M)^{\eta+1}A^{\eta+1}-W_M M \tag{28}$$

$$\implies \Xi_i\left(\eta+1\right)F^{\eta}F_M A^{\eta+1}=W_M \tag{29}$$

$$\implies PA=\frac{W_M}{\left(\eta+1\right)F_M}, \tag{30}$$

Or:

$$\frac{\partial}{\partial m}\mathbb{E}\left[p+a|\ell,k,m\right]=\frac{\partial}{\partial m}\mathbb{E}\left[w_m|\ell,k,m\right]-\frac{\partial}{\partial m}\log F_M \tag{31}$$

Suppose there are decreasing returns to materials, $F_{MM}<0$. Then,

$$\frac{\partial}{\partial m}\log F_M=\frac{F_{MM}}{F_M}\times M\leq 0\implies\frac{\partial}{\partial m}\mathbb{E}\left[p+a|\ell,k,m\right]\geq\frac{\partial}{\partial m}\mathbb{E}\left[w_m|\ell,k,m\right]. \tag{32}$$

Supposing that W_M does not vary across plants — or, at least, does not vary conditional on (ℓ,k) because (ℓ,k) proxy for that variation — then,

$$\frac{\partial}{\partial m}\mathbb{E}\left[p+a|\ell,k,m\right]\geq 0 \tag{33}$$

$$\implies\frac{\partial}{\partial m}\mathbb{E}\left[p+q|\ell,k,m\right]\geq\frac{\partial f}{\partial m}. \tag{34}$$

Otherwise if W_M does vary and the variation is not fully proxied for by (ℓ,k), then, intuitively, $\mathbb{E}\left[w_m|\ell,k,m\right]$ would be decreasing in materials which might make the bounds slightly too tight.

To limit the bias in the upper bound caused by unobserved wage variation, we could condition on additional variables besides (ℓ,k,m). For example, we could include plant level and time level fixed effects in the regression of revenue on (ℓ,k,m). In that case, the identification assumption would be that there exists an x such that,

$$\frac{\partial}{\partial m}\mathbb{E}\left[w|\ell,k,m,x\right]\approx 0\implies\frac{\partial}{\partial m}\mathbb{E}\left[p+q|\ell,k,m,x\right]\geq\frac{\partial f}{\partial m}. \tag{35}$$

Or, in the most ideal case, if W_M is observed (but output prices are not), then we can use the above argument to write,

$$\frac{\partial}{\partial m}\mathbb{E}\left[p+q|\ell,k,m\right]-\frac{\partial}{\partial m}\mathbb{E}\left[w_m|\ell,k,m\right]\geq\frac{\partial f}{\partial m}, \tag{36}$$

and increase the credibility of our bounds with revenue data as output.

5 Inference

We now show how to use the identification assumptions to make inference on statistics of the markup distribution.

In Appendix A, we show how nonparametric inference could be made on the production function and other partially-identified statistics of interest. While those techniques are general, put few assumptions on the data generating process, and can be used to test arbitrary hypothesis, they are also often difficult to use in practice, requiring us to make choices about tuning parameters and to search high-dimensional parameter spaces.

Our preferred empirical approach is to assume the output elasticity of the flexible input is linear in a finite number of parameters,

$$\frac{\partial f}{\partial m}(\ell,k,m)=\sum_{j=1}^{J}\theta_j r_j(\ell,k,m), \tag{37}$$

Where $r_j(\cdot)$ are known basis functions.

For any given material output elasticity (a vector of parameters θ), we can recover the full production function via the Markov restrictions on productivity as shown in Gandhi, Navarro, and Rivers (2015). We first establish that the production function identified in this way is a linear function of θ as well, allowing us to build a computationally-convenient estimator.

Let $\overline{m}$ be a fixed value of material use. Then, we can recover the production function up to a function of labor and capital with knowledge of the material output elasticity,

$$\int_{\overline{m}}^{m}\frac{\partial f}{\partial m}(\ell,k,m)\,dm=f(\ell,k,m)-f\left(\ell,k,\overline{m}\right). \tag{38}$$

Because productivity is,

$$a_t=q_t-f(\ell_t,k_t,m_t)-\epsilon_t=q_t-\int_{\overline{m}}^{m_t}\frac{\partial f}{\partial m}(\ell_t,k_t,m)\,dm-f\left(\ell_t,k_t,\overline{m}\right), \tag{39}$$

the Markov assumption on productivity and the assumption that labor and capital are quasi-fixed, uncorrelated with e_t, identifies $f\left(\ell_t,k_t,\overline{m}\right)$,

$$a_t=g(\ell_{t-1},k_{t-1},m_{t-1})+e_t \tag{40}$$

$$\implies q_t-\int_{\overline{m}}^{m_t}\frac{\partial f}{\partial m}(\ell_t,k_t,m)\,dm=f\left(\ell_t,k_t,\overline{m}\right)+g(\ell_{t-1},k_{t-1},m_{t-1})+e_t+\epsilon_t \tag{41}$$

$$\mathbb{E}\left[e_t+\epsilon_t|\ell_t,k_t,\ell_{t-1},k_{t-1},m_{t-1}\right]=0. \tag{42}$$

With $f\left(\ell_t, k_t, \overline{m}\right)$ in hand, we can identify the full production function, $f(\ell, k, m)$, from equation (38).

Empirically, we approximate the functions $f\left(\ell, k, \overline{m}\right)$ and $g(\ell, k, m)$ with flexible functional forms,

$$f\left(\ell, k, \overline{m}\right) = s(\ell, k)^{\top} \gamma, \quad g(\ell, k, m) = p(\ell, k, m)^{\top} \rho, \tag{43}$$

where γ and ρ are unknown parameter vectors and $s(\cdot)$ and $p(\cdot)$ are known functions.

After a little algebra, we can write γ as a linear function of θ,

$$q_t - \sum_{j=1}^{J} \theta_j \int r_j(\ell_t, k_t, m)\, dm = s(\ell_t, k_t)^{\top} \gamma + p(\ell_{t-1}, k_{t-1}, m_{t-1})^{\top} \rho + e_t + \epsilon_t \tag{44}$$

$$\implies \begin{pmatrix} \gamma \\ \rho \end{pmatrix} = \mathbb{E}\left[(s,p)^{\top}(s,p)\right]^{-1} \mathbb{E}\left[(s,p)^{\top}\left(q_t - \sum_{j=1}^{J} \theta_j \int r_j(\ell_t, k_t, m)\, dm\right)\right] \tag{45}$$

$$\gamma = \Gamma_0 + \Gamma_1 \theta, \tag{46}$$

where Γ_0 and Γ_1 are matrices with elements that are point identified from the data (functions of the moments of s, p, and q) and θ are the parameters of the flexible output elasticity.

The full production function is then a linear function of θ,

$$f(\ell, k, m) = s^{\top}\Gamma_0 + s^{\top}\Gamma_1\theta + \sum_{j=1}^{J} \theta_j \int_{\overline{m}}^{m} r_j\, dm. \tag{47}$$

So because θ is partially identified, the material output elasticity, from our assumptions in Section 3, we have partial identification of the full production function as well.

The object of interest, a linear function of the regression coefficients of markups regressed on some plant characteristics X, is also a linear function of θ because

$$\frac{P}{\text{MC}} = X^{\top}\beta + U \tag{48}$$

$$\beta = B\theta, \tag{49}$$

and we want upper and lower bounds on some element of β given that θ is in the set identified by our assumptions.

Our identified set for θ is,

$$\exp\left\{\mathbb{E}\left[\log \frac{W_M M_t}{\text{Revenue}_t \times \mathbb{E}\left[\exp\left(q_t - \mathbb{E}\left[q_t | \ell_t, k_t, m_t\right]\right)\right]} | \ell_t, k_t, m_t\right]\right\} \tag{50}$$

$$\leq r(\ell_t, k_t, m_t)^{\top} \theta \leq \frac{\partial}{\partial m_t} \mathbb{E}\left[q_t | \ell_t, k_t, m_t\right] \tag{51}$$

$$\forall (\ell_t, k_t, m_t). \tag{52}$$

We approximate both the lower and upper bound functions by some functional form. It is good in practice to ensure that the upper and lower bounds are about as flexible as we allow the materials output elasticity to be. Otherwise, the bounds may be narrow in finite samples because of our functional form assumptions and not because of variation in the data or the strength of our identifying assumptions. It is difficult to find a place for a line, say, between two very jagged curves.

So, say the bounds are, after choosing our functional forms (this choice is discussed more in the following sections),

$$r(\ell_t, k_t, m_t)^\top \underline{\theta} \leq r(\ell_t, k_t, m_t)^\top \theta \leq r(\ell_t, k_t, m_t)^\top \overline{\theta}, \tag{53}$$

where $\underline{\theta}$ and $\overline{\theta}$ can be estimated by $\sqrt{n}$- asymptotically normal estimators.

We can write the hypothesis that, say, $\beta_1 = b$, as the hypothesis that there exists a θ such that,

$$r(\ell_{it}, k_{it}, m_{it})^\top \left(\underline{\theta} - \theta\right) \leq 0 \tag{54}$$

$$r(\ell_{it}, k_{it}, m_{it})^\top \left(\theta - \overline{\theta}\right) \leq 0 \tag{55}$$

$$B_1^\top \theta - b = 0. \tag{56}$$

We take the view that factor usage is fixed (so the r above are not random) as is commonly done to compute standard errors in ordinary least squares regressions. Conditional on the "design" of the model, we can apply a convenient inference strategy based on a standard Wald hypothesis test.

Write the above constraint set as,

$$M\theta \leq d - (b, -b, 0, \ldots, 0)^\top, \tag{57}$$

where b is the hypothesized value of $B_1^\top \theta$ and the linear equality constraint has been transformed into two inequality constraints.

For any given θ we can test the hypothesis that $B_1^\top \theta = b$ by performing a simple Wald test[3],

$$\min_{t \geq 0} \left(\widehat{d} - \widehat{M}\theta - (b, -b, 0, \ldots, 0)^\top\right)^\top \widehat{W}(\theta, b) \left(\widehat{d} - \widehat{M}\theta - (b, -b, 0, \ldots, 0)^\top\right), \tag{58}$$

Where $\widehat{W}$ is the inverse of the covariance matrix of $\widehat{d} - \widehat{M}\theta - (b, -b, 0, \ldots, 0)^\top$. Kudo (1963) shows that the limiting distribution of the above test statistic is a mixture of chi-squared distributions (the weights of the mixture can be easily approximated via Monte Carlo techniques).

We then search over θ to find a value of θ where the hypothesis that $B_1^\top \theta = b$ is not rejected to form a confidence interval for $B_1^\top \theta$.

We use the following steps to test whether $B_1^\top \theta = b$.

[3] It is possible to use Guggenberger, Hahn, and Kim (2008) to improve the power of this test. Guggenberger, Hahn, and Kim (2008) show that a solution exists to a linear inequality if and only if a specially constructed vector is entirely positive so applying the Wald test to this transformed model has more power. But, it turns out, the vector is very large in our particular problem.

1. For a given hypothesis $B_1^\top \theta = b$ and θ, solve the following optimization problem to obtain $\mathrm{TS}(\theta)$:

$$\mathrm{TS}(\theta) = \min_{t \geq 0} \left(\widehat{d} - \widehat{M}\theta - (b, -b, 0, \ldots, 0)^\top\right)^\top \widehat{W}(\theta, b) \left(\widehat{d} - \widehat{M}\theta - (b, -b, 0, \ldots, 0)^\top\right) \quad (59)$$

2. Draw N_{sim} random vectors Z with distribution $N\left(0, \widehat{W}^{-1}\right)$ and for each $ns = 1, \ldots, N_{sim}$, solve the following quadratic program,

$$t_{ns} = \arg\min_{t \geq 0} (Z_{ns} - t)^\top \widehat{W} (Z_{ns} - t) \quad (60)$$

   Let $s_{ns} = \sum_{j=1}^{J} \mathbf{1}\left(t_{ns,j} = 0\right)$ where J is the number of rows of $\widehat{C}$ (and the dimension of t). For each $j = 0, \ldots, J$,

$$\omega_j = \frac{1}{N_{sim}} \sum_{ns=1}^{N_{sim}} \mathbf{1}\left(s_{ns} = j\right) \quad (61)$$

   The distribution of the test statistic under the null is then,

$$\mathrm{pr}\{TS \leq u\} = \sum_{j=0}^{J} \omega_j \mathrm{pr}\left\{\chi_j^2 \leq u\right\}, \quad (62)$$

   where χ_j^2 is the chi-squared distribution with j degrees of freedom and χ_0^2 is a point mass on zero.

3. Use the distribution to compute the p-value of the hypothesis that $B_1^\top \theta = b$ and $\theta = \theta$. Then, search across θ to find a value of θ where the p-value is greater than the threshold required for the confidence interval.

We then repeat this for each b to construct a confidence interval for the parameter.

5.1 A specific estimator to use in practice

In this section, we give a "cookbook" version of our estimator we found works well in practice.

Step 1. Choose how flexible to allow the materials output elasticity to be. The translog production function has a linear-in-logs output elasticity, a good place to start.

Step 2. Given our choice of a flexible materials elasticity, estimate the bounds using the same functional form. Here, we will assume the output elasticity is chosen to be linear-in-logs.

Step 3. Estimate the upper bound on the materials elasticity by regressing log output on the translog function of log labor, log capital, and log materials, which gives a linear-in-log inputs upper bound to the materials' elasticity,

$$q_{it} = \rho_\ell \ell_{it} + \rho_{\ell\ell}\ell_{it}^2 + \rho_{\ell k}\ell_{it}k_{it} + \rho_{\ell m}\ell_{it}m_{it} + \rho_k k_{it} + \rho_{km}k_{it}m_{it} + \rho_{kk}k_{it}^2 + \rho_m m_{it} + \rho_{mm}m_{it}^2 + \upsilon_{it}^q. \tag{63}$$

Step 4. Estimate a linear-in-log inputs approximation to the lower bound, fitting the following non-linear model,

$$\log\left[\frac{W_M M_t}{\text{Revenue}_t \times \mathbb{E}\left[\exp(\epsilon_t)\right]}\right] = \log\left[\kappa_0 + \kappa_\ell \ell_t + \kappa_k k_t + \kappa_m m_t\right] + \upsilon_t^r \tag{64}$$

$$\mathbb{E}\left[\upsilon_t^r | \ell_t, k_t, m_t\right] = 0. \tag{65}$$

In levels, the model is linear so we can estimate it with OLS (it has a multiplicative residual but is otherwise standard),

$$\frac{W_M M_t}{\text{Revenue}_t \times \mathbb{E}\left[\exp(\epsilon_t)\right]} = \left[\kappa_0 + \kappa_\ell \ell_t + \kappa_k k_t + \kappa_m m_t\right]\exp\left(\upsilon_t^r\right) \tag{66}$$

Step 5. With estimates of κ and ρ in hand, we form the linear inequalities that define the identified set,

$$\kappa_0 + \kappa_\ell \ell_{it} + \kappa_k k_{it} + \kappa_m m_{it} \leq \theta_0 + \theta_\ell \ell_{it} + \theta_k k_{it} + \theta_m m_{it} \leq \rho_m + \rho_{\ell m}\ell_{it} + \rho_{km}k_{it} + 2\rho_{mm}m_{it}. \tag{67}$$

Step 6. Recall that we want to know bounds on a linear function of coefficients from a linear regression of markups (P/MC) on some plant characteristics X,

$$\frac{P}{MC} = X^\top \beta + E, \quad \mathbb{E}[XE] = 0 \tag{68}$$

$$\text{Statistic of interest} = \tau = c^\top \beta. \tag{69}$$

Because P/MC is,

$$\frac{P}{MC} = \frac{Revenue}{W_M M} \times (\theta_0 + \theta_\ell \ell + \theta_k k + \theta_m m), \tag{70}$$

a linear function of θ, to compute the linear function of θ that corresponds to the statistic of interest, we can write,

$$\frac{Revenue_{it}}{W_M M_{it}} y_{it} = X^\top \beta_y + E_{it}^y \quad \mathbb{E}\left[XE_{it}^y\right] = 0 \tag{71}$$

$$\tau_y = c^\top \beta_y \quad \text{for all } y \in \{1, \ell, k, m\} \tag{72}$$

$$\tau = \tau_1\theta_0 + \tau_\ell\theta_\ell + \tau_k\theta_k + \tau_m\theta_m \tag{73}$$

Where $(\tau_1, \tau_\ell, \tau_k, \tau_m)$ can all be estimated directly from the data.

Step 7. Use the hypothesis test to form confidence intervals for τ.

If we are interested in a statistic of productivity or of the full production function, then we also need to choose functional forms for $f\left(k_t, \ell_t, \overline{m}_t\right)$ and $g\left(\ell_{t-1}, k_{t-1}, m_{t-1}\right)$. We suggest using translog functional forms for both.

6 Constant returns to scale

It is possible to achieve point identification in the standard proxy structure of Section 2 without assuming conduct. For example, constant returns to scale plus the proxy structure point identifies the production function. The basic intuition is:

1. From Gandhi, Navarro, and Rivers (2015), we know that for a given materials output elasticity (f_m), we can recover the production function.
2. From Constant Returns to Scale, we know that $f_m = 1 - f_k - f_\ell$. This creates a fixed point problem which, we will show (under some assumptions) has a unique solution.

While the structure of constant returns to scale is significant, point identification (or, at least, narrower bounds) may be necessary in contexts where we want to rule out perfect competition (we want to find evidence of market power or to find how average markups have changed over time). The bounds we presented in earlier sections are useful for discovering which kinds of firms or plants have more or less market power.

Write:

$$\frac{\partial f}{\partial m} := h(\ell, k, m) \tag{74}$$

$$q_t - \int_{\overline{m}}^{m_t} h(\ell_t, k_t, m)\, dm = f\left(\ell_t, k_t, \overline{m}\right) + g(\ell_{t-1}, k_{t-1}, m_{t-1}) + \eta_t \tag{75}$$

$$\implies \mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] = f\left(\ell_t, k_t, \overline{m}\right) + g(\ell_{t-1}, k_{t-1}, m_{t-1}) \tag{76}$$

Because,

$$f(\ell, k, m) = f\left(\ell, k, \overline{m}\right) + \int_{\overline{m}}^{m} h(\ell, k, m)\, dm \tag{77}$$

$$\implies \frac{\partial f}{\partial \ell} = \frac{\partial f}{\partial \ell}\left(\ell, k, \overline{m}\right) + \int_{\overline{m}}^{m} \frac{\partial h}{\partial \ell}(\ell, k, m)\, dm \tag{78}$$

$$\frac{\partial f}{\partial k} = \frac{\partial f}{\partial k}\left(\ell, k, \overline{m}\right) + \int_{\overline{m}}^{m} \frac{\partial h}{\partial k}(\ell, k, m)\, dm, \tag{79}$$

constant returns to scale implies,

$$\frac{\partial f}{\partial k}\left(\ell, k, \overline{m}\right) + \int_{\overline{m}}^{m} \frac{\partial h}{\partial k}(\ell, k, m)\, dm + \frac{\partial f}{\partial \ell}\left(\ell, k, \overline{m}\right) + \int_{\overline{m}}^{m} \frac{\partial h}{\partial \ell}(\ell, k, m)\, dm + h(\ell, k, m) = 1. \tag{80}$$

We can then substitute expressions for the other terms in terms of the data and the unknown function h,

$$\frac{\partial}{\partial k_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] + \int_{\overline{m}}^{m} \frac{\partial h}{\partial k}(\ell, k, m)\, dm \tag{81}$$

$$+\frac{\partial}{\partial \ell_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] + \int_{\overline{m}}^{m} \frac{\partial h}{\partial \ell}(\ell, k, m)\, dm + h(\ell, k, m) = 1. \tag{82}$$

If h is linear-in-parameters (θ) then the problem has the form,

$$c(\ell_{t-1}, k_{t-1}, m_{t-1}, k_t, \ell_t)^\top \theta + b(\ell_{t-1}, k_{t-1}, m_{t-1}, k_t, \ell_t) = 1, \tag{83}$$

where c and b are functions point-identified by the data. Then, for example,

$$\mathbb{E}\left[cc^\top\right]\theta = \mathbb{E}\left[c(1-b)\right] \tag{84}$$

If $\mathbb{E}\left[cc^\top\right]$ is invertible, then θ is identified. Other identification conditions would work too (for example, the condition that there exist at least J linearly independent vectors $c(\ell_{t-1}, k_{t-1}, m_{t-1}, k_t, \ell_t)$).

For more nonparametric results, suppose that there are two functions h_0 and h_1 that satisfy the conditions. Then:

$$\begin{aligned}
h_1(\ell, k, m) - h_0(\ell, k, m) &= \frac{\partial}{\partial k_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h_0(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] + \int_{\overline{m}}^{m} \frac{\partial h_0}{\partial k}(\ell, k, m)\, dm \\
&+\frac{\partial}{\partial \ell_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h_0(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] + \int_{\overline{m}}^{m} \frac{\partial h_0}{\partial \ell}(\ell, k, m)\, dm \\
&-\frac{\partial}{\partial k_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h_1(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] - \int_{\overline{m}}^{m} \frac{\partial h_1}{\partial k}(\ell, k, m)\, dm \\
&-\frac{\partial}{\partial \ell_t}\mathbb{E}\left[q_t - \int_{\overline{m}}^{m_t} h_1(\ell_t, k_t, m)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] - \int_{\overline{m}}^{m} \frac{\partial h_1}{\partial \ell}(\ell, k, m)\, dm \\
= \frac{\partial}{\partial k_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} (h_1 - h_0)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] &+ \frac{\partial}{\partial \ell_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} (h_1 - h_0)\, dm | \ell_t, k_t, \ell_{t-1}, k_{t-1}, m_{t-1}\right] \\
&- \int_{\overline{m}}^{m} \frac{\partial}{\partial k}(h_1 - h_0)\, dm - \int_{\overline{m}}^{m} \frac{\partial}{\partial \ell}(h_1 - h_0)\, dm
\end{aligned}$$

If h does not vary with m, then we have nonparametric identification.

The identification result can be made nonparametric at a high level by making the following definitions and assumptons:

1. h can be approximated by a linear-in-parameter sieve,

$$\inf_{\theta, J} \sup_{\ell, k, m} |h(\ell, k, m) - \sum_{j=1}^{J} r_j(\ell, k, m)\theta_j| = 0 \tag{85}$$

2. Define c_j as,

$$c_j(\ell,k,m,\ell_{t-1},k_{t-1}) = -\frac{\partial}{\partial k_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} r_j(\ell_t,k_t,m')\,dm'|\ell_t,k_t,\ell_{t-1},k_{t-1},m_{t-1}\right] + \int_{\overline{m}}^{m}\frac{\partial r_j}{\partial k}(\ell,k,m')\,dm'$$
$$-\frac{\partial}{\partial \ell_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} r_j(\ell_t,k_t,m')\,dm'|\ell_t,k_t,\ell_{t-1},k_{t-1},m_{t-1}\right] + \int_{\overline{m}}^{m}\frac{\partial r_j}{\partial \ell}(\ell,k,m')\,dm' + r_j(\ell,k,m)$$

3. Suppose that $\mathbb{E}\left[c_J c_J^\top\right]$ has an eigenvalue bounded away from 0 across J.
4. Define the sequence $h_J = r_J^\top \theta_J$. $\lim h_J = h$.

We present results under both our weaker, partial identification conditions and using the above point identification results under the constant returns to scale assumption.

7 Nonparametric identification

Let f and h be two homogenous of degree one functions on (ℓ,k,m) and define $z \equiv f_m - h_m$. Let the space of functions z be called $\mathcal{Z}$. The following condition is necessary and sufficient for identification,

$$\left\{z \equiv \frac{\partial}{\partial k_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} z\,dm|\ell_t,k_t\right] + \frac{\partial}{\partial \ell_t}\mathbb{E}\left[\int_{\overline{m}}^{m_t} z\,dm|\ell_t,k_t\right] - \int_{\overline{m}}^{m}\frac{\partial z}{\partial k}dm - \int_{\overline{m}}^{m}\frac{\partial z}{\partial \ell}dm\right\} \tag{86}$$
$$\implies z \equiv 0 \text{ or } z \not\in \mathcal{Z} \tag{87}$$

Proof of sufficiency. Trivial.

Proof of necessity. Suppose the identifiation assumption were not true

8 Markups and trade

Having identified the main policy parameter of interest and proposed a method to estimate it, we now turn to the specifics of the problem we face as well as a simple Monte Carlo to show the dangers of ignoring the identification problems we raised in Section 2 when attempting to answer the main question.

We are interested in understanding the differences in markups between plants who export and plants who do not. We start with exactly how we measure markups in the data given that plants may belong to two markets and their markups can be different in each market. We then give a simple theoretical model that illustrates the main predictions of the theory we will test empirically and use that model as the basis for a Monte Carlo illustrating how our identification method leads to correct inference about the relationship between markups and export status when current methods in the literature would not.

8.1 Defining markups

Let λ be the the Lagrange multiplier for the output constraint in the cost minimization problem, W_M be the price of materials, and F be $\exp(f)$, the level production function. Because plants minimize their variable costs and materials are chosen flexibly, the first order conditions for the cost minimization problem give,

$$W_M = \lambda \frac{\partial F}{\partial M} A \iff W_M M = \lambda \frac{\partial f}{\partial m} Q \tag{88}$$

$$\implies \lambda = \frac{W_M M}{Q} \times \left[\frac{\partial f}{\partial m} \right]^{-1}, \tag{89}$$

and, by the envelope theorem, marginal cost is equal to λ.

To recover markups, suppose a plant gets one price for the goods it sells in the domestic market (P_{dom}) and another price in the world market (P_{world}),

$$Q = Q_{\text{dom}} + Q_{\text{world}} \tag{90}$$

$$\frac{P_{\text{dom}}}{\text{MC}} \times \frac{Q_{\text{dom}}}{Q} + \frac{P_{\text{world}}}{\text{MC}} \times \frac{Q_{\text{world}}}{Q} = \frac{\text{Revenue}}{W_M M} \times \frac{\partial f}{\partial m}, \tag{91}$$

where Q_{dom} is output sold to the domestic market and Q_{world} is output sold to the world market.

Because we observe revenue, we can recover the left hand side if we can identify the output elasticity for materials.

For non-exporters, the left hand side is their price-to-marginal cost ratio, a natural measure of markups. For exporters, the measure is an output-share weighted measure of markups in the domestic market and markups in the world market. When estimating the relationship between export status and markups, we will need to control for the amount a plant exports relative to its total output or else differences between world price and domestic price might cause us to make misleading inferences about the validity of the theory that exporters have higher markups *within their home market*[4]. The risk is not hypothetical: because, we will show, Chilean exporters earn lower markups in the world market than domestically, not controlling for the relative intensity with which a plant competes in the world market versus the domestic market, would lead us to conclude export status is negatively related to markups but we usually find support for the opposite conclusion when we control for how much of a plant's total output goes to the world versus domestic markets.

8.2 A simple model where greater productivity gives greater markups

In this section, we use a simple model to show the main theoretical prediction we test: does the greater productivity of exporters give them greater market power? The model is a simplified

[4]Ideally, we would be able to use domestic markups and world markups separately but we do not observe a distinction between materials used to produce products to export and materials used to produce products for the domestic market.

version of Melitz and Ottaviano (2008) and a broad class of other monopolistic competition models with firm heterogeneity.

Each firm i produces its own unique variety. The inverse demand for variety i is

$$P_i = \alpha - \beta Q_i - \eta \int Q_i di = \alpha - \beta Q_i - \eta \overline{Q}. \tag{92}$$

Each demand parameter (α, β, η) is positive. There is a continuum of small firms so each takes $\overline{Q}$ as given when maximizing their profit. Each firm has a constant marginal and average cost C_i.

In the closed economy, when there is only one market, the firm's profit maximization problem gives,

$$\max_{Q_i} \left(\alpha - C_i - \beta Q_i - \eta Q\right) Q_i \tag{93}$$

$$\implies Q_i = \frac{\alpha - C_i}{2\beta} - \frac{\eta}{2\beta}\overline{Q}. \tag{94}$$

Markups are then (ignoring that when markups would be negative by this formula, the firm will choose zero output),

$$\alpha + \frac{C_i - \alpha}{2} + \frac{\eta}{2}\overline{Q} - \eta \overline{Q} - C_i = \frac{\alpha}{2} - \frac{C_i}{2} - \frac{\eta}{2}\overline{Q}, \tag{95}$$

and so markups are clearly decreasing in C_i; firms with greater costs (lower productivity), have lower markups.

Now suppose that there is a second market where inverse demand is,

$$P_i' = \alpha' - \beta' Q_i' - \eta' \int Q_i' di = \alpha' - \beta' Q_i' - \eta' \overline{Q}', \tag{96}$$

and each demand parameter is positive.

A domestic firm can enter the second market at a cost, σ. Because all firms are small, when they decide which market to enter, they will not consider how they effect the total mass of firms within each market. Let $\overline{Q}$ be the total output of firms in the domestic market and $\overline{Q}'$ be the total output in the foreign market. A firm in the domestic market solves the market entry problem,

$$\max\left\{\Pi\left(C_i; \overline{Q}\right) + \Pi'\left(C_i; \overline{Q}'\right) - \sigma, \Pi\left(C_i; \overline{Q}\right)\right\}, \tag{97}$$

where $\Pi(\cdot;\cdot)$ is the profit function in the domestic market and $\Pi'(\cdot;\cdot)$ is the profit function in the foreign market.

So domestic firms enter the foreign market if and only if,

$$\Pi'\left(C_i; \overline{Q}'\right) \geq \sigma. \tag{98}$$

Because $\Pi'\left(\cdot;\overline{Q}'\right)$ is a strictly decreasing function for all $\overline{Q}'$, exporters are the firms with,

$$C_i \leq \left(\Pi'\right)^{-1}\left(\sigma;\overline{Q}'\right). \tag{99}$$

So, for any market equilibrium $\left(\overline{Q},\overline{Q}'\right)$, firms with lower costs will be exporters. So, within each market, exporters will be the more productive firms and enjoy higher markups.

8.3 Monte Carlo

We demonstrate with a Monte Carlo how misleading our inferences about markups can be if we ignore the identification problem raised in Section 2. The model is similar to the model in the previous section in that it has a model of monopolistic competition in mind so there is some concavity in the plant's revenue function. In order to give the proxy model the best chance of doing well, we have one dynamic input (capital) that is fixed in period t so that it does not suffer from the flexible input identification problem.

We generate the data by treating each plant as if it were its own monopoly — without entry and exit, there is nothing fundamentally different between this model and a monopolistically competitive model.

We use a simple model where k_t (capital) is the quasi-fixed input and materials m_t is the flexible input (used as the proxy variable). There is no labor input. The firm solves the following Bellman equation:

$$\begin{aligned} V(A,K) = \max_{M,X}\left[M^{\theta}K^{\gamma}A\right]^{\eta+1} - W_M M - W_K K - W_X X + \beta\mathbb{E}\left[V(Av,\delta K + X)\right] \\ \log v \sim N\left(0,\sigma^2\right) \end{aligned}$$

with parameter values:

$$\begin{aligned} \theta = 0.6, \gamma = 0.3, \sigma = 0.5, \beta = 0.95, \delta = 0.9, W_K = 0.5, W_X = 0.2, \\ \log(K_0 - 1) \sim N(0,1), \log A_0 \sim N(0,1) \\ W_M \sim U[0.3, 0.5], \eta = -0.4. \end{aligned}$$

Because the revenue function is strictly concave, the firm will earn positive markups. We allow the materials input price to vary across plants in order to demonstrate:

1. Even with observed variation in input prices, the De Loecker and Warzynski (2012) approach based on the proxy model will fail to estimate the correct markups because, conditional on lagged wages, the second lag of materials has no power as an instrument.
2. Second, even though our identification approach does not explicitly allow for unobserved variation in input prices, our bounds are fairly robust to this variation.

The statistic of interest is the average difference in (P/MC) between a group we will call "exporters" and a group we will call "non-exporters" in anticipation of our application. In constructing the Monte Carlo, "export" status will simply be a dummy indicating the plant likely has higher markups,

$$\text{“Exporter”} = \mathbf{1}\left(\frac{P}{MC} - 2 + \text{shock} > 0\right) \quad \text{shock} \sim N(0,1).n \tag{100}$$

The De Loecker and Warzynski (2012)-based estimator we use assumes the production function is Cobb Douglas (as it is), uses a translog functional form to approximate $\mathbb{E}\left[q_{it}|m_{it}, k_{it}, w_{m,it}\right]$, and correctly assumes an AR1 process for productivity. Because materials appears in the proxy function, we use the second lag of materials as an instrument, and because capital is a fixed input, we use k_{it} as an instrument because it is uncorrelated with the innovation shock. Because, with knowledge of w_{it-1}, there is no variation in w_{it} that is correlated with m_{it-2} (there is no variation in w_{it} at all after conditioning on w_{it-1}), m_{it-2} will not have any power as an instrument.

For the bounds results, We use the exact estimator we proposed in Section 5.1 without taking advantage of the fact that the true production function is Cobb-Douglas[5].

In Table 1, we give the results of the 10,000 simulation Monte Carlo. The key takeaways are,

1. The density of the proxy estimates is bi-modal because it is not identified (see Figure 1). 17.23% of the time the estimate of the difference in markups is *negative* even though, by construction, the indicator for export status is based purely on whether the firm has high markups. While the bias of the estimator is not so bad (see Table1), the bi-modal nature of the estimator makes that a misleading way to judge the estimate.

2. In constrast, our lower bound on the effect of export status on markups is never negative so we would always make the correct inference about the direction of the average difference in markups.

3. Our bounds contain the true value of the average difference in markups 91.55% of the time (these bounds are not confidence intervals).

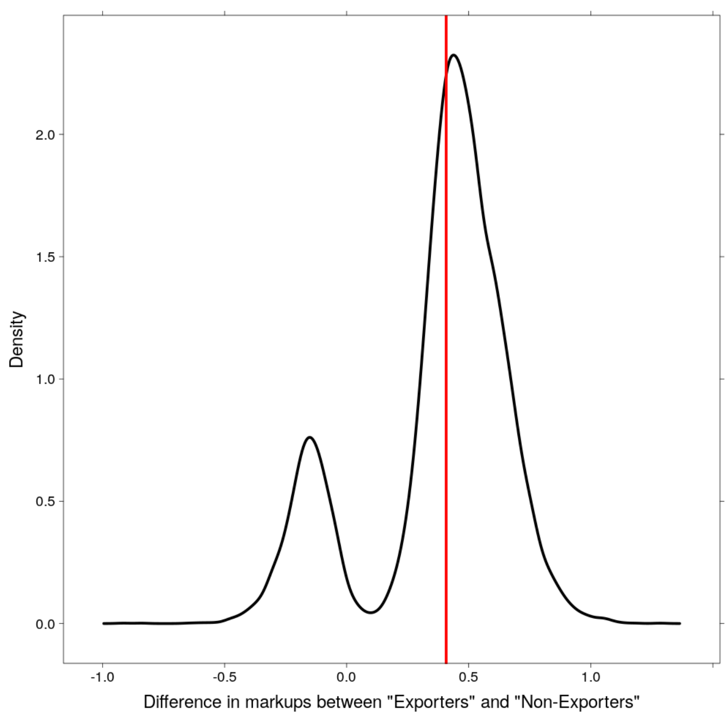
4. The upper bound estimate is weak for this particular data generating process (mean value of upper bound on difference ≈ 2.7; true value of difference is ≈ 0.4), but the lower bound estimate (the Gandhi, Navarro, and Rivers 2015 estimator) is tighter.

[5]If we knew the production function was Cobb-Douglas, there would be little point in worrying about identifying markups for a regression problem because the regression of *log* markups ($\log P - \log MC$) on covariates would be identified without estimating the flexible output elasticity at all.

Table 1: Monte Carlo Results

Estimator	Statistic	Value
True Model	True Value	0.408
	Mean Estimate	0.379
Proxy	Standard deviation	0.290
	Probability negative	0.172
	Mean Estimate	$[0.395, 2.738]$
Bounds	Standard deviation	$\{0.024, 0.516\}$
	Probability negative	0.000
	Probability bounds contain estimate	0.916

Figure 1: Proxy estimates of the difference in markups between exporters and non-exporters (Monte Carlo); red line is true difference



9 Data

We apply our identification strategy to study the relationship between export status and markups, using a plant level dataset from Chile's Instituto Nacional de Estadística's census of manufacturing plants with more than 10 employees. The dataset has been used in several other papers in the production function estimation and trade literature, see Levinsohn and Petrin (2003), Pavcnik (2002), and Gandhi, Navarro, and Rivers (2015).

Our measure of plant output is deflated revenue. If there is only one output price in each industry, then it is fine to treat revenue as output because the price will just change the location of the log production function—which is not identified anyway. But there may be problems if there is price variation that the deflator can not pick up—of course, this is a long-standing issue with production data and not unique to our data and problem.

We measure the labor input by number of employees, capital using data on investment and depreciating that invest to recover capital (as in Greenstreet 2007), and deflated spending on materials as our measure of material use.

The panel runs from 1990 to 1996.

We define the industries at the two-digit level (each industry will have its own production function, f).

In Table 2, we give descriptive statistics of the industries that are important for understanding competition within the industry which is our ultimate object of interest. The table contains information about the industry's size (total sales), measures of competition within the industry (sales concentration and the number of plants), and the number of exporters and how that number has changed over time. Generally, the number of exporters has increased over time while plant concentration measures have remained relatively stable (with the exception of the Chemicals industry).

Table 2: Descriptive statistics of industries

Industry	Year	Sales (1985 Trillion Pesos)	C8	Plants	Exporters
Pooled	1990	6.3	27.0%	3625	666
	1996	13.9	13.7%	4059	1013
Food	1990	1.6	18.7%	1142	215
	1996	4.7	19.6%	1154	294
Textiles	1990	0.4	19.4%	703	91
	1996	0.9	22.2%	661	160
Wood products	1990	0.2	28.2%	330	72
	1996	0.8	30.6%	455	103
Paper	1990	0.4	59.0%	194	32
	1996	1.2	50.3%	224	61
Chemicals	1990	1.2	58.2%	444	128
	1996	1.9	25.1%	531	190
Non-metal minerals (not oil/coal)	1990	0.2	64.2%	122	22
	1996	0.6	54.3%	162	28
Basic metals	1990	1.7	75.0%	60	24
	1996	2.4	70.4%	61	30
Fabricated metal products	1990	0.6	26.2%	586	76
	1996	1.4	26.1%	727	132

Notes. C8 is the eight-plant sales concentration (sales of the eight plants with greatest sales divided by total industry sales).

10 Do exporters have higher revenue-to-cost ratios?

Before turning to our main results, we first consider what we can learn from the data and reduced-form measures of the effect of the export status on markups, both to illustrate the advantages of modeling markups in a structural way and to see what we can learn from the data before arguing for a certain structural interpretation of the data.

Suppose the industry production function had constant returns to scale, each input was chosen flexibly, and plants minimized costs. Then,

$$\frac{P}{\text{MC}} = \frac{PQ}{W_L L + W_K K + W_M M}, \tag{101}$$

price over marginal cost is equal to the ratio of revenue-to-costs[6].

More generally, the ratio of revenue-to-costs contains information on two things: (1) the markup of the plant and (2) the productivity of the plant[7]. So establishing the relationship between plant export status and the revenue-to-costs ratio establishes how export status predicts the combination of markups and productivity. In the following section, we will use our partial identification strategy to parse out how export status is related to markups and productivity, separately. But first, we establish the combined effect from the data.

We run the following regression,

$$\left(\frac{PQ}{W_L L + W_K K + W_M M}\right)_{ijt} = \tilde{\tau} \times \text{Exporter}_{ijt} + \beta_k k_{ijt} + \beta_\ell \ell_{ijt} + \beta_e \times \frac{\text{Exports}_{ijt}}{\text{Revenue}_{ijt}} + \mu_t + \xi_j + u_{ijt}, \tag{102}$$

where i indexes plant, j indexes the 4-digit industry the plant belongs to, and t indexes year[8]. We run the regression 2-digit industry-by-industry and also pooling across industries. The regression is similar to De Loecker and Warzynski (2012)'s model but with the notable difference that we include the ratio of exports-to-revenue on the right hand side so that $\tilde{\tau}$ is the difference between the *domestic* revenue-to-cost ratio of exporters and the domestic revenue-to-cost ratio of non-exporters. This control is necessary because theory only predicts that exporters will have higher markups within their own market—not that they will have higher markups in the foreign market. In the following sections, we will replace the ratio of revenue to input expenditures with the ratio of price to marginal cost, but the form of the regression will be the same.

In most industries $\tilde{\tau}$ is positive, exporters have higher revenue-to-cost ratios (see Table 3), giving reduced form evidence for the prediction that exporters earn higher markups. But another

[6]This result is true because the materials output elasticity is then equal to $W_M M/(W_L L + W_K K + W_M M)$.

[7]The returns to scale matter too, but by assuming the industry production function depends only on input use, we have already assumed the returns to scale are the same for exporters and non-exporters. Of course, reality might be different.

[8]Capital is difficult to measure. We define capital as deflated cumulative investment as in Greenstreet (2007). Nominal capital is the not deflated version of the measure with is what we take $W_K K$ to be. The other input prices are clearer.

Table 3: Relationship between Export Status and the Revenue-to-Cost Ratio

Industry	Estimate	Standard Error
Pooled	5.2%	0.8%
Food	7.1%	1.8%
Textiles	3.9%	1.6%
Wood products	1.0%	2.2%
Paper	−1.3%	2.8%
Chemicals	2.5%	1.6%
Non-metal minerals (not oil/coal)	16.3%	4.4%
Basic metals	11.1%	5.2%
Fabricated metal products	9.9%	2.1%

explanation for the higher revenue-to-cost ratios has not been ruled out by these regressions: if exporters are more productive, we might observe the same result even if the exporters have no market power. We need a structural measure of markups in order to conclude that exporters had greater market power.

11 Markups and export status

Many trade models predict exporters will have higher markups than non-exporters because they predict more productive firms will have higher markups and that exporters are more productive, like the model in Section 8.2, Melitz and Ottaviano (2008), and Bernard, Eaton, Jensen, and Kortum (2003). Whether the prediction is true matters for the welfare effect of trade liberalization because the theory predicts that, when a country opens up to trade, the domestic firms that engage in trade will be the firms which also have more market power.

De Loecker and Warzynski (2012) also study whether exporters have higher markups empirically, and they find support for this prediction. Relative to their work, aside from our contributions to the methodology of recovering markups, we allow for the relationship between markups and export status to vary by industry and control for the possibility that exporters enjoy different markups in foreign markets. But the basic form of our regressions is similar to theirs to isolate the effect of our new identification strategy.

We also make a small contribution to the large literature that establishes the positive relationship between many different measures of productivity and whether a plant exports by confirming the positive relationship holds in the industries we study—and because we use weaker assumptions to identify productivity than the rest of the structural productivity literature, the result is even more robust. See the citations in Syverson (2011)'s survey for a list of papers that find exporters are more productive.

Let μ be our measure of markups,

$$\mu = \frac{\text{Revenue}}{W_M M} \times \frac{\partial f}{\partial m}. \tag{103}$$

Our main regression model is,

$$\mu_{ijt} = \tau \times \text{Exporter}_{ijt} + \beta_\ell \ell_{ijt} + \beta_k k_{ijt} + \beta_e \times \frac{\text{Exports}_{ijt}}{\text{Revenue}_{ijt}} + \delta_t + \xi_j + u_{ijt}, \tag{104}$$

Where τ is the coefficient of interest and j indexes the 4-digit industry so we have industry fixed effects at the 4-digit level. τ is the difference in markups between exporters and non-exporters when the percentage of exported revenue is zero, which allows us to compare the domestic markups of exporters with the domestic markups of non-exporters. The theory in Section 8.2 predicts $\tau \geq 0$.

Recall our measure of markups is,

$$\mu = \frac{Q_{\text{dom}}}{Q} \times \left(\frac{P}{MC}\right)_{dom} + \frac{Q_{\text{world}}}{Q} \times \left(\frac{P}{MC}\right)_{world}, \tag{105}$$

which is why we need to control for exports to make sure we are comparing domestic markups to domestic markups (otherwise, exporters might have lower markups in the world market than in the domestic market and we would pick up a negative relationship between export status and markups but that would not violate the theory in Section 8.2). Theory is less clear on the sign of β_e; it is an empirical question whether exporters have higher markups in Chile or in the world, an empirical question we answer. But, intuitively, competition is tougher when we include all the world's competitors than among only Chilean competitors, implying the sign is negative.

We bound τ (and β_e) industry-by-industry at the two-digit level. We also present results that pool the data and bound τ using data from all industries (i.e. force the production function to be the same for all industries).

We find that, generally, exporters have higher domestic markups in Chile than non-exporters as predicted in Section 8.2. While there are some industries where the theory does not hold (the Paper and Chemicals industries), the higher markup result holds when we pool across industries and in most industries when we approach the problem industry-by-industry.

We also find strong evidence that $\beta_e < 0$—that is, Chilean exporters have lower markups in the foreign market than they do domestically (only in the Paper industry is $\beta_e > 0$). The world market is more competitive than the domestic market.

We let the material output elasticity be a linear function of log input use,

$$\frac{\partial f}{\partial m} = \theta_0 + \theta_\ell \ell + \theta_k k + \theta_m m, \tag{106}$$

and use the functional forms we proposed in Section 5.1.

To test the theory, we need to bound the coefficients in two regressions,

$$\mu_{ijt} = \tau \times \text{Exporter}_{ijt} + \beta_\ell \ell_{ijt} + \beta_k k_{ijt} + \beta_e \times \frac{\text{Exports}_{ijt}}{\text{Revenue}_{ijt}} + \delta_t + \xi_j + u_{ijt} \tag{107}$$

$$a_{ijt} = \tau_a \times \text{Exporter}_{ijt} + \beta_{a,\ell} \ell_{ijt} + \beta_{a,k} k_{ijt} + \delta_{a,t} + u_{a,ijt} \tag{108}$$

The theories we want to test correspond to the following sign restrictions on the parameters,

(1) Exporters have higher markups $\implies \tau \geq 0$.

(2) Exporters are more productive $\implies \tau_a \geq 0$.

(3) If exporters are more productive, they have higher markups $\implies \tau\tau_a \geq 0$.

We find strong support for the prediction that exporters are more productive. In the pooled results, $1.1\% \leq \tau_a \leq 7.3\%$, and in most industries, exporters are more productive than non-exporters of similar size (only in the Wood Products industry is there evidence that exporters are less productive than non-exporters). We also find support for the prediction that exporters have higher markups. In the pooled results, $-0.9\% \leq \tau \leq 2.8\%$ but in most industries the bounds are positive. But in the Paper and Chemicals industry there is evidence that exporters have lower markups than non-exporters. Because we find τ_a mostly positive and τ mostly positive, we find evidence for the third prediction as well: that more productive plants tend to have higher markups. But, in the Wood Products industry, we find $\tau_a < 0$ and $\tau > 0$, contradicting the prediction.

See Table 4 for full results, including bounds on β_e, all of which are negative, implying exporters have lower markups in the world market than domestically. This result is robust and very well-identified by the economic assumptions we use for identifiecation.

We do not observe wage variation so we can not use De Loecker and Warzynski (2012)'s method to identify the production function as we argued in Section 2. But to see what we would get, in the results presented in Table 5, we ignore the non-identification result and use the second lag of log material use as an instrument. Most of the estimates from using this method are outside the bounds we constructed using our method.

But that the De Loecker and Warzynski (2012) method often produces "reasonable" estimates suggests there are issues we need to overcome with the proxy model more generally. The non-identification problem with using lagged inputs is not that the instruments are *invalid* (an untestable hypothesis) but that the instruments have no *power* (a testable hypothesis). To estimate the production function using the De Loecker and Warzynski (2012) approach (we do not observe wages), we approximate $f(\ell_{it}, k_{it}, m_{it})$ and $g(\ell_{it-1}, k_{it-1}, m_{it-1})$ by second order polynomials and use the natural excluded instruments

$$\left(m_{it-2}, m_{it-2}\ell_{it}, m_{it-2}k_{it}, m_{it-2}^2\right)$$

(everywhere m_{it} appears in $f(\cdot)$ replace it with m_{it-2}).

In the proxy model without wage variation (the model in Section 2 and Gandhi, Navarro, and Rivers 2015), m_{it-2} and its various interactions are valid instruments (they are uncorrelated with the residual) but they have no power because m_{it-2} can not predict e_{it} — the shock to productivity — which is the only separate variation in m_{it}, given $(\ell_{it}, k_{it}, \ell_{it-1}, k_{it-1}, m_{it-1})$ (see Section 2). We can test whether m_{it-2} has any power as an instrument using a Cragg and Donald (1993) test which tests the null hypothesis that the model is underidentified. If the test is rejected, then the proxy model without wage variation is rejected because the instrument should be weak.

We can see from Table 6 that the test is rejected at conventional significance levels in most industries and in the pooled sample. This result suggests that the proxy model is missing something: innovations in productivity are not just noise but depend on other plant characteristics that are proxied by m_{it-2}. There are two potential responses to this result: (1) get more columns in the dataset, allow the production function to vary by more plant characteristics, or (2) do not require the innovations to be random noise, allow there to be unobserved plant-level variables that affect productivity, so that this result does not contradict the model—see, for example, the partial identification approach in Flynn (2018).

Lastly, a bit of caution. While the particular goal of this paper is to show how the proxy assumptions can identify markups, the idea that markups have a distribution can present a unique challenge to the proxy model's assumption that there is only one unobserved state variable, productivity. Because markups can only vary across state variables, this restricts the extent to which markups can vary across plants. We might imagine that some market power is derived from product differentiation and, if we want to allow for these additional unobservable state variables to affect the markup distribution, we need to replace the invertibility assumption with another assumption that bounds the flexible output elasticity from above but allows for multiple unobserved state variables. The "linear positive association" from Flynn (2018) satisfies both of these requirements and can be used as an alternative to the invertibility assumption to bound the flexible output elasticity from above. If we impose the linear positive association assumption, we also do not need the Markov structure to achieve meaningful identification of productivity because linear positive association restricts all output elasticities—but the Markov assumption does provide substantially more identifying power (at the cost of being somewhat ad-hoc).

Table 4: Relationship between Export Status, Productivity, and Markups

Industry	Dependent Variable	LB (90%)	LB	UB	UB (90%)	CRS
Pooled	Markups		−0.9%	2.8%		−1.2%
	Productivity		1.1%	7.3%		4.9%
	β_e		−28.0%	−18.2%		−24.3%
Food	Markups		2.7%	6.2%		1.6%
	Productivity		−5.0%	4.7%		−4.7%
	β_e		−30.4%	−21.4%		−35.9%
Textiles	Markups		2.7%	3.3%		7.6%
	Productivity		6.6%	10.1%		5.3%
	β_e		−37.7%	−28.0%		−27.7%
Wood products	Markups		6.9%	10.2%		1.4%
	Productivity		−18.7%	−9.4%		−8.4%
	β_e		−46.1%	−30.2%		−8.1%
Paper	Markups		−9.9%	−4.1%		−110.4%
	Productivity		8.4%	14.7%		−32.9%
	β_e		6.5%	16.2%		−36.5%
Chemicals	Markups		−15.4%	−11.9%		−41.9%
	Productivity		−7.2%	1.8%		−14.6%
	β_e		−49.3%	−34.2%		−51.6%
Minerals	Markups		6.5%	9.1%		10.5%
	Productivity		5.0%	9.5%		1.7%
	β_e		−41.1%	−25.5%		−66.2%
Basic metals	Markups		15.1%	22.7%		17.8%
	Productivity		6.4%	16.3%		20.2%
	β_e		−42.1%	−21.1%		−25.4%
Fabricated metal products	Markups		1.9%	7.8%		0.8%
	Productivity		9.0%	14.6%		12.4%
	β_e		−33.4%	−21.1%		−19.8%

Notes. LB (90%) and UB (90%) are the lower and upper limits of the 90% confidence interval. CRS estimates are the point estimates from assuming constant returns to scale.

Table 5: Relationship between Export Status, Productivity, and Markups (using the second lag of log materials as an instrument)

Industry	Dependent Variable	Estimate
Pooled	Markups	15.2%
	Productivity	0.1%
Food	Markups	8.0%
	Productivity	−13.0%
Textiles	Markups	11.5%
	Productivity	2.0%
Wood products	Markups	8.8%
	Productivity	−15.4%
Paper	Markups	$-(13.2\times 100)\,\%$
	Productivity	−3.2%
Chemicals	Markups	$-(1.0\times 100)\,\%$
	Productivity	−19.0%
Minerals	Markups	18.9%
	Productivity	6.9%
Basic metals	Markups	54.3%
	Productivity	0.1%
Fabricated metal products	Markups	16.9%
	Productivity	1.3%

Table 6: Cragg and Donald Test of Power of m_{it-2}

Industry	Test Statistic	P-value
Pooled	80.1	0.0%
Food	20.4	0.0%
Textiles	8.5	0.4%
Wood products	7.0	0.8%
Paper	0.2	96.8%
Chemicals	0.9	34.4%
Non-metal minerals (not oil/coal)	9.6	0.2%
Basic metals	0.6	45.4%
Fabricated metal products	25.1	0.0%

12 Conclusion

Identifying markups by directly recovering marginal costs can allow us to recover markups while putting less structure on conduct and to learn about competition in industries with poor demand data. But identifying marginal costs faces difficult simultaneity problems that are often overcome in the traditional production function and cost function estimation literature by making the conduct assumptions we are trying to avoid. We show that the standard proxy model can partially identify the production function, and so, partially identify marginal costs, without making these assumptions.

References

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

Andrews, D. and X. Shi (2014). Nonparametric inference based on conditional moment inequalities. *Journal of Econometrics 179*, 31–45.

Andrews, D. and X. Shi (2017). Inference based on many conditional moment inequalities. *Journal of Econometrics 196*(2), 275–287.

Bernard, A., J. Eaton, J. B. Jensen, and S. Kortum (2003). Plants and productivity in international trade. *American Economic Review 93*(4), 1268–1290.

Chetverikov, D. (2013). Testing regression monotonicity in econometric models. *Working Paper*.

Cragg, J. and S. Donald (1993). Testing identifiability and specification in instrumental variable models. *Econometric Theory 9*(2), 222–240.

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71.

Flynn, Z. (2018). Identifying productivity when it is a choice. *Working Paper*.

Gandhi, A., S. Navarro, and D. Rivers (2015). On the identification of production functions: how heterogeneous is productivity? *Working Paper*.

Greenstreet, D. (2007). Exploiting sequential learning to estimate establishment-level productivity dynamics and decision rules. In *Economic Series Working Papers*, Number 345. University of Oxford.

Guggenberger, P., J. Hahn, and K. Kim (2008). Specification testing under moment inequalities. *Economics Letters 99*, 375–378.

Kudo, A. (1963). A multivariate analogue of the one-sided test. *Biometrika 50*, 403–418.

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

Marschak, J. and W. Andrews (1944). Random simulataneous equations and the theory of production. *Econometrica 12*, 143–205.

Melitz, M. and G. Ottaviano (2008). Market size, trade, and productivity. *Review of Economic Studies 75*, 295–316.

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

Pavcnik, N. (2002). Trade liberalization, exit, and productivity improvements: evidence from chilean plants. *Review of Economic Studies 69*, 245–276.

Syverson, C. (2011). What determines productivity? *Journal of Economic Literature 49*(2), 326–365.

A Nonparametric inference

We first show how nonparametric inference on the production function itself might be done given the two identifying assumptions. We will not use this method in our application where we will instead take advantage of the form of the final statistic we are interested in — a coefficient from the regression of markups on plant characteristics, not the production function itself — and use a flexible parametric form for the production function. We include a method of nonparametric inference on the production function for completeness and to connect our identifying assumptions to inference problems previously studied in the econometrics literature.

The inequality restrictions can be written as,

$$\frac{\partial}{\partial m_t}\mathbb{E}\left[q_t - f(\ell_t, k_t, m_t)\,|\,\ell_t, k_t, m_t\right] \geq 0 \tag{109}$$

$$\mathbb{E}\left[\log\frac{\partial f}{\partial m_t} - \log\left(\frac{W_M M}{\text{Revenue}\times\mathbb{E}\left[\exp\left(q_t - \mathbb{E}\left[q_t|\ell_t, k_t, m_t\right]\right)\right]}\right)|\ell, k, m\right] \geq 0. \tag{110}$$

Under the null hypothesis that $f = f_0$, the first identifying assumption is a test that a nonparametric regression of $q_t - f(\ell_t, k_t, m_t)$ on (ℓ_t, k_t, m_t) is increasing in m_t. Chetverikov (2013) presents a hypothesis test that can be used to test whether nonparametric regressions are increasing in a given variable.

The second identifying assumption is a conditional moment inequality. Andrews and Shi (2014) develop a method of nonparametric inference on conditional moment inequalities.

The tests can be combined using a valid, but likely conservative, method of inference using the Bonoferri bounds.

The following is a valid nonparametric test of the hypothesis that $f = f_0$ given our identifying assumptions:

1. Compute the p-value, π_1, of the Chetverikov (2013) test that the first identifying assumption is true.
2. Compute the p-value, π_2, of the Andrews and Shi (2014) test that the second identifying assumption is true.
3. Reject the null hypothesis if either $\pi_1 \le \alpha/2$ or $\pi_2 \le \alpha/2$ where α is the significance level.

If we want to be relatively more conservative with one of the identification assumptions, the confidence set will have the same size if we reject the null if either $\pi_1 \le \omega\alpha$ or $\pi_2 \le (1-\omega)\,\alpha$ for some $0 < \omega < 1$.

An alternative approach that may have more power is to use Andrews and Shi (2017)'s method of inference on (un)countably many conditional moment inequalities because the first identification assumption can be written as uncountably many conditional moment inequalities and so can the second (as a finite number of conditional moment inequalities).