Measuring Markups with Production Data

Zach Flynn, Amit Gandhi, and James Traina*

Afiniti, University of Pennsylvania, and University of Chicago

June 13, 2019

Abstract

We show standard methods to estimate production functions do not identify markups. This nonidentification creates spurious skewness in estimated markup distributions. We also show that ex-ante structure on the returns to scale solves the identification problem. In US public firm data and in a Monte Carlo experiment, we find that applying constant returns to scale performs remarkably well and reduces the skewness in the markup distribution among public-firm by as much as half in comparison to nonidentified estimates. This results in half the efficiency losses in output and labor shares when calibrated to a recent macroeconomic model.

Keywords: Market Power, Markups, Production Functions, Productivity, Labor Share
JEL Classification: E2, L11, L13, D24

*We thank Enghin Atalay, John Grigsby, Ken Hendricks, Ivan Kirov, Alex Sollaci, and Alan Sorensen for helpful discussions. We're grateful to Virgiliu Midrigan and Michael Peters for their correspondence on markups in macroeconomic models. We also thank seminar participants at Chicago Booth's Stigler Center Workshop, the ASSA 2019 Meetings, and UChicago's Industrial Organization Workshop for their comments and feedback. Corresponding author address: traina@uchicago.edu.

An emerging class of macroeconomic research argues that the rise of firm market power explains several important secular trends in the global economy: among them are falling output growth and labor shares (Karabarbounis and Neiman (2013)), and rising industrial concentration and corporate profits (Barkai (2017)).[1] A key mechanism that connects these empirical patterns with the rise of market power is the growing importance of very large (mega) or highly productive (superstar) firms with very high markups.

In this paper, we empirically study whether "mega markups" have emerged in the US economy of the last 40 years. We first show that the standard proxy variables techniques for measuring markups with production data, which underlie many current studies, do not identify markups. This nonidentification creates spurious skewness in the estimated markup distribution.

However, we also show markups *are* identified if we add further structure to the production function. We specifically study the identifying power of knowing the returns to scale (RTS). If the RTS is known ex ante, or the RTS depends on fixed or dynamic inputs (e.g. capital), then we show both the production function and firm markups are separately identified.

When external information on the RTS is unavailable, a practical way to apply our approach is to impose constant returns to scale (CRS). A variety of applied production papers (e.g. Basu and Fernald (1997), Syverson (2004), Foster et al. (2008), and Bloom et al. (2012)) argue CRS is a good approximation. CRS has also become standard in much applied work with firm-level data as well as macroeconomic models with heterogeneous firms (Atkeson and Burstein (2008); Melitz and Ottaviano (2008); Hsieh and Klenow (2009); Asker et al. (2014); Peters (2018); Edmond et al. (2018); Eggertsson et al. (2018)). Our identification argument gives rise to a natural extension of the standard GMM estimators, which makes the RTS structure simple to use in practice. For illustration, we present the moment conditions explicitly for the Cobb-Douglas and translog production functions.

We apply our approach to estimate production functions for US public firms and document significant advantages of our estimator against real data. The standard existing approach to markup estimation emanates from De Loecker and Warzynski (2012), which combines insights from Hall (1988) with proxy variable production function methods.[2] We call this combination the DLW estimator.[3] We find that the DLW estimator produces markups that are similar to the theoretical upper bound implied by the proxy structure alone (i.e. without assuming CRS), exceeding the upper bound for more than 40% of firm-year observations. Consequently, underidentified estimators are a key driver of previous estimates of mega markup firms. Moreover, these markups are unstable across common specification choices: under a Cobb-Douglas specification, the DLW estimator produces aggregate markup estimates just above 1; under a translog specification, this statistic jumps to above 6. As a reference point from a product with famously inelastic demand, the cigarette smoker demand elasticities in Chaloupka (1991) under pure monopoly imply a markup of 2. Our proposed

[1] See Syverson (2019) for a broad review of this literature.

[2] By far the most popular estimators in this vein are Olley and Pakes (1996), Levinsohn and Petrin (2003), and Ackerberg et al. (2015). As of June 2019, these papers collectively have over 12,000 Google Scholar citations.

[3] For example, see De Loecker and Eeckhout (2017).

identification strategy estimates a lower variance and skewness of markups and produces stable estimates across specifications.

We explore the economic implications of our markup estimates by appealing to the recent literature on markups in macroeconomic models.[4] In standard models, efficient allocation of resources implies low and relatively equalized markups. High markup firms should see increased resource flows, decreasing their markups. To quantify the efficiency implications of our empirical results, we rely on sufficient statistics proposed in the Peters (2018) model of innovation, firm dynamics, and heterogeneous markups. The stationary markup distribution is Pareto, whose tail index measures efficiency losses from lower creative destruction. Fitting our baseline estimates to a Pareto distribution and hence the Peters model, we find annual output losses from markups of about 1 percent of GDP (100B to 200B USD annually). Furthermore, the labor share is 11 percent lower than an idealized benchmark. These estimates are similar in size to the findings of Edmond, Midrigan, and Xu (2018), which uses an alternative heterogeneous firm model that is not directly calibrated to the empirical markup distribution. Estimates from the DLW model would incorrectly *double* these losses.

We also explore the implications of our identification strategy through a quantitatively calibrated Monte Carlo analysis. We assume the data generation process follows a workhorse dynamic macro model parametrized closely to the empirical calibration of Edmond, Midrigan, and Xu (2018). In our Monte Carlo experiments, we find that the bias from ignoring the flexible input identification problem is large. Nonidentification produces biases that are *twenty times* as large in percentage terms as our proposed estimator when the true RTS is one, and *four times* as large even when the true RTS is drawn uniformly between [0.9, 1.1].

Controlling the data generation process allows us to investigate an alternative source of identification suggested in the literature: serially correlated input prices that vary across firms. Such variation overcomes the flexible input problem when: (1) firm-level prices are observable, and (2) the input price variation is orthogonal to productivity and output prices. These conditions are largely impractical. (1) is rarely available in production datasets that span the number of industries required to inform macroeconomic research questions. (2) means the price variation cannot come from quality differences – firms would need to pay higher prices for inputs because they are unlucky, not because they are purchasing better inputs or because of changes in demand or conduct conditions. In our Monte Carlo, even when these prices are completely orthogonal to productivity, the DLW estimator's bias is still roughly *four times* the bias of our CRS estimator when the prices are unobserved.

Our paper contributes to the recent literature on market power in the macroeconomy. DLE uses the DLW estimator to argue that the average markup has risen precipitously since 1980, concentrated in the upper tail of mega markup firms. Under our CRS assumption, our analysis indicates that half of this upper tail skewness is the spurious outcome of nonidentification. Our finding helps rationalize the differences between the firm-level markup estimates of DLE and the industry-level markup estimates of Hall. The latter paper updates the Hall

[4] These models evolved from recent work in international economics that quantifies markup dispersion and misallocation in response to trade shocks. They do so by specifying parametric models of consumer demand and firm conduct. See Epifani and Gancia (2011), Edmond, Midrigan, and Xu (2015), Edmond et al. (2018), and Peters (2018).

1988 analysis to find a modest rise in markups driven by finance, utilities, and healthcare, with little relation to the share of mega-firms within industries. We would like to emphasize that low skewness in markups is not inconsistent with high skewness in size or productivity (**?**). For example, in the Melitz and Ottaviano (2008) model of heterogeneous firms, larger and more productive firms charge lower markups. This feature is an endogenous outcome of competitive pressures dominating a selection effect of lower marginal costs.

Our paper also contributes to the industrial organization literature on identifying technology. Identifying the flexible input elasticity is notoriously difficult: Both traditional and modern production function estimators suffer from a classic simultaneity problem called *transmission bias*.[5] The same optimization conditions used to recover markups imply firms endogenously choose flexible inputs depending on their productivity. Consequently, when flexible input expenditure is high, we cannot disentangle whether it is because the output elasticity is large, or because the firm is very productive. Gandhi, Navarro, and Rivers (2017) (GNR) formalize this intuition in a general setting by proving that the elasticity is not identified under the standard restrictions in the proxy variable literature. Because identifying markups relies on identifying the elasticity, markups are not identified. We offer a starting point for solving this fundamental problem by reducing it to questions on the RTS of production.

Section 1 reviews the proxy variable production model and the identification problem of the flexible input elasticity. Sections 2 presents our solution in the Cobb-Douglas and general nonparametric cases. In Section 3, we offer explicit estimation algorithms for the commonly used Cobb-Douglas and translog production specifications. Sections 4 and 5 compare our estimator with the standard DLW estimator in Monte Carlo and public-firm data. We discuss extensions using variable RTS and partial identification restrictions in Section 6. Section 7 concludes with suggestions for future econometric research on measuring markups.

1 The Inconsistency of the Proxy Model

In this section, we demonstrate that the proxy model is not identified under the assumptions required to estimate markups. But more importantly, we find a way out by pinpointing the exact source of underidentification: the returns to scale is not identified. This opens a path forward to resolve the identification problem that we pursue in the later sections of this paper.

Recovering identification of the production approach to markup estimation is particularly important given the research agenda on market power in the macroeconomy. This approach is agnostic about consumer demand and firm conduct, and is therefore compatible with a large class of models with imperfect competition. Moreover, it scales well to multi-industry settings where the demand approach to markup estimation is infeasible because of data requirements. Finally, this approach allows us to study firm heterogeneity, which is critical for understanding the underlying mechanisms of production.

Let q_t be log output, and (v_t, k_t) be the log of flexible inputs and capital in period t. Denote

[5] Marschak and Andrews (1944) first articulated this problem in a general setting. Hoch (1958) developed it further for Cobb-Douglas production functions, coining the "transmit" terminology.

the level of these variables by their respective capitalizations. Flexible inputs v_t are variable and static: they contribute to time t's production, but have no direct effect on the firm's future decisions. Capital k_t is fixed and dynamic: it can not be adjusted in period t (the firm chooses it in an earlier period), but it affects period t output (and potentially output in other periods).[6] Consequently, the firm treats the capital stock k_t as a state variable at time t (chosen before productivity innovations are known to the firm). The data are inputs and output over the panel $t = 1, \ldots T$,

$$\{(q_t, v_t, k_t)\}_{t=1}^{T} \tag{1}$$

The joint distribution of the data in the underlying population of firms is identified in the data.

The proxy approach to production function estimation relates these data to a model of production that consists of three parts. We label this set of assumptions the *proxy structure*:

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

$$q_t = f\left(v_t, k_t\right) + a_t + \epsilon_t \tag{2}$$

   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, so that $\mathbb{E}\left[\epsilon_t | v_t, k_t\right] = 0$.

2. Productivity follows a first-order Markov process:

$$a_t = g\left(a_{t-1}\right) + \eta_t, \tag{3}$$

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

$$\mathbb{E}\left[\eta_t | k_t, v_{t-1}, k_{t-1}, \ldots\right] = 0, \tag{4}$$

   where the ellipses represent all other lags of the inputs.

3. Flexible input demand has the form $v_t = v\left(a_t, k_t\right)$ where $v(\cdot, k_t)$ is strictly increasing in a_t.[7]

The model is called the proxy structure because this last assumption implies productivity can be proxied by observable inputs, $a_t = v^{-1}\left(v_t, k_t\right) = a(v_t, k_t)$.

1.1 Nonidentification

Gandhi, Navarro, and Rivers (2017) show that 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

[6] More generally, k_t can be a vector of inputs each of which is not flexible, i.e. either fixed, dynamic, or both.

[7] For details on how these assumptions change with deflated sales data, see Appendix A.

of observables. Estimators using only these restrictions are inconsistent, making the resulting estimates difficult to interpret.

GNR also shows that if the flexible input elasticity $f_v(v_t, k_t)$ were identified, then the production function f would be identified over the support of the data given the proxy structure. The source of nonidentification is the flexible input elasticity, or as we'll soon prove is equivalent, the returns to scale (the sum of the flexible input elasticity and the capital elasticity).

The envelope theorem applied to the variable cost minimization problem implies that marginal cost is

$$MC = \frac{WV}{Q} \times \frac{1}{f_v} \implies \mu = \frac{P}{MC} = \frac{PQ}{WV} \times f_v \tag{5}$$

where W is the price of the flexible input V. GNR uses the relationship (5) combined with assuming the markup μ to identify the production function (they assume $P = MC$ in their main specification). If the markup is known, then the flexible input elasticities can be recovered, and this identifies the remaining parameters of the production function under the proxy structure.

The DLW estimator is remarkably powerful in its generality,[8] but it also prevents us from using the GNR solution to the GNR problem. We cannot use the GNR solution because it assumes the markup which is our object of interest. We need another structural assumption to identify the output elasticity of the flexible input, and therefore the markup.

1.2 Partial Identification

While the proxy structure alone does not point identify the production function in the presence of flexible inputs, Flynn and Gandhi (2018) show that it does partially identify the production function by offering lower and upper bounds on the flexible input elasticity. In our empirical analysis, we use these bounds as evaluation criteria for our CRS estimator against the DLW estimator.

The markup upper bound under the proxy structure is the markup we would compute if we estimated the production function via OLS, ignoring endogeneity. The proxy structure tells us that flexible input demand is strictly increasing in productivity: $a_t = v^{-1}(v_t, k_t) = a(v_t, k_t)$, and $a_v > 0$. As a consequence, the inverse flexible input demand is increasing as well, so that $\frac{\partial}{\partial v_t}\mathbb{E}[q_t|v_t, k_t] = f_v + a_v > f_v$. This condition offers an upper bound on the flexible input elasticity, and so an upper bound on markups: $\frac{PQ}{WV} \times \frac{\partial}{\partial v_t}\mathbb{E}[q_t|v_t, k_t] > \frac{PQ}{WV} \times f_v = \mu$. Ignoring transmission bias by estimating the production function via OLS will return estimates that are biased strictly upwards under the proxy structure

The markup lower bound under the proxy structure is even simpler: markups can never fall below 1. The assumption that the relevant input is variable and chosen without dynamic

[8] While the proxy assumption was originally developed in a model of price-taking firms, Flynn (2019) shows that it holds so long as firms face a demand function in which MR(Q)*Q is an increasing function of Q, which implies the profit maximization problem with zero costs has no interior solution. Among many other models, this condition is true for perfect competition; or monopolistic competition facing constant elasticity or logit demand.

consequence rules out typical theoretical justifications for markups below 1, such as dynamic pricing. In other words, to measure markups from the firm's first order condition as in Hall (1988), we must assume away mechanisms that cause markups to fall below 1. This lower bound is a useful diagnostic for estimation output. For example, Hall (2018) evaluates the original Hall (1988) method in updated data by interpreting markups below 1 as sampling error.

2 Nonparametric Identification with Returns to Scale

We resolve the same identification problem addressed in Gandhi, Navarro, and Rivers (2017), but instead of assuming we know the markup, we assume that we know the returns to scale of the production function. We assume a returns to scale of 1 (i.e. CRS) both for ease of exposition, and because that is what we use in our empirical application. However, all arguments follow for any known returns to scale. We extend our results in Section 6 to allow for unknown variation in RTS through fixed inputs.

CRS identifies the production function in the following way:

1. For each flexible input elasticity, there exists a unique production function that both has that elasticity and satisfies the proxy assumptions (GNR).
2. If the nonflexible input elasticities are known, then the flexible input elasticity is identified under CRS because RTS is the sum of the output elasticities. For each set of nonflexible input elasticities, there exists a unique production function.

Consequently, we have a fixed point problem in f_v:

$$f_v\left(v,k\right) = 1 - f_k\left(v,k;f_v\right) \tag{6}$$

We show that the fixed point problem has a unique solution.

Theorem 1 presents our general, nonparametric identification result. Aside from a known returns to scale, we must also make a technical assumption about the distribution of flexible inputs v_t conditional on current fixed inputs k_t. The assumption requires that the parts of v_t that are not collinear with k_t are correlated with k_t. For example, identification fails if $v_t = k_t + \varepsilon_t$, where ε_t is uncorrelated with k_t.

Theorem 1. If, for any function $\Delta\left(v,k\right)$ such that for almost all (v,k),

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]\right]+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta\left(v',k\right)dv'+\Delta\left(v,k\right)=0 \tag{7}$$

it must be the case that $\int|\Delta\left(v,k\right)|=0$. Then the proxy structure implies either:

1. There exists no production function in the set identified by the proxy structure with constant returns to scale.

2. For any two production functions with constant returns to scale in the identified set, f^1 and f^0,

$$\int_v \int_k |\left(f^1 - f^0\right)(v,k)|dkdv = 0. \tag{8}$$

Proof. For proof, see Appendix A. □

We can use Theorem 1 to generate identification conditions which are easy-to-check (and empirically testable) for particular functional form choices for the production function. To do so, replace Δ with the functional form of f_v to recover what the identification assumption means for that functional form.

The following result is useful for understanding what the identification condition means in the context of particular functional forms:

Result 1. If $\Delta(v,k)$ satisfies,

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t} \Delta(v', k_t)\, dv' | k_t = k, v_{t-1}, k_{t-1}\right]\right] + \int_{\overline{v}}^{v} \frac{\partial}{\partial k}\Delta(v', k)\, dv' + \Delta(v,k) = 0, \tag{9}$$

then it must be the case that Δ is homogenous of degree zero (regardless of data generating process), i.e.

$$\Delta_v + \Delta_k = 0 \tag{10}$$

Proof. Differentiate the identification condition with respect to v. □

This result is useful for narrowing down the kind of production functions we have to consider as potentially part of the identified set. We present more explicit identifying conditions derived from the general condition above for two common functional form choices: Cobb-Douglas and translog production functions. Both make use of Result 1.

2.1 Illustrating Identification with the Cobb-Douglas Production Function

For illustration, consider identification for Cobb-Douglas production functions. When the production function is Cobb-Douglas, we have the following equations from the proxy structure introduced above,

$$q_t = \theta_v v_t + \theta_k k_t + a_t + \epsilon_t \tag{11}$$

$$a_t = g(a_{t-1}) + \eta_t \tag{12}$$

$$\mathbb{E}[\eta_t | k_t, v_{t-1}, k_{t-1}, \ldots] = 0 \tag{13}$$

$$\mathbb{E}[\epsilon_t | v_t, k_t] = 0 \tag{14}$$

$$a_t = a(v_t, k_t) \tag{15}$$

Constant returns to scale removes one parameter from the problem so the number of parameters is equal to the number of instruments (the only instrument with power is capital). We can write the production equation as

$$q_t = (1 - \theta_v) k_t + \theta_v v_t + a_t + \epsilon_t \tag{16}$$

$$\implies q_t - k_t = \theta_v (v_t - k_t) + g(a_{t-1}) + \eta_t + \epsilon_t \tag{17}$$

$$= \theta_v (v_t - k_t) + g(a(k_{t-1}, v_{t-1})) + \eta_t + \epsilon_t \tag{18}$$

$$\tag{19}$$

Applying the moment restrictions we have from the proxy structure, we have

$$\mathbb{E}[q_t - k_t | k_t, v_{t-1}, k_{t-1}] = \theta_v (\mathbb{E}[v_t | k_t, v_{t-1}, k_{t-1}] - k_t) + g(a(k_{t-1}, v_{t-1})),$$

The production function is identified if Assumption 1 holds:

Assumption 1. There exists (k_t, k_{t-1}, v_{t-1}) such that,

$$\frac{\partial}{\partial k_t} \mathbb{E}[v_t | k_t, k_{t-1}, v_{t-1}] \neq 1. \tag{20}$$

Choose a (k_t, k_{t-1}, v_{t-1}) such that $\partial \mathbb{E}[v_t | k_t, k_{t-1}, v_{t-1}] / \partial k_t \neq 1$. Then,

$$\frac{\partial}{\partial k_t} \mathbb{E}[q_t - k_t | k_t, v_{t-1}, k_{t-1}] = \theta_v \left(\frac{\partial}{\partial k_t} \mathbb{E}[v_t | k_t, k_{t-1}, v_{t-1}] - 1 \right), \tag{21}$$

and θ_v is identified. Assumption 1 is exactly equivalent to the general identification condition of Theorem 1 if Δ is a constant, as it is in the Cobb Douglas production function.

2.2 Illustrating Identification with the Translog Production Function

In the translog functional form, $\Delta(v, k) = \delta_0 + \delta_v v + \delta_k k$. The identification condition is that, there does not exist $\delta \neq 0$ such that

$$\delta_0 \times \left\{ -\mathbb{E}\left[\frac{\partial}{\partial k} \mathbb{E}[v | k_t = k, v_{t-1}, k_{t-1}] \right] + 1 \right\} \tag{22}$$

$$+\delta_v \times \left\{ -\frac{1}{2} \mathbb{E}\left[\frac{\partial}{\partial k} \mathbb{E}\left[v_t^2 | k_t = k, v_{t-1}, k_{t-1} \right] \right] + v \right\} \tag{23}$$

$$+\delta_k \times \left\{ -\mathbb{E}\left[\frac{\partial}{\partial k} \left(k \mathbb{E}[v - \overline{v} | k_t = k, v_{t-1}, k_{t-1}] \right) \right] + (v - \overline{v}) + k \right\} = 0 \tag{24}$$

$$\delta_0 u_0 + \delta_v u_v + \delta_k u_k = 0 \tag{25}$$

If there exist at least three (v, k) such that the three vectors (u_0, u_v, u_k) are linearly independent, then the only δ that satisfies the equation is $\delta = 0$. A sufficient condition is:

$$\mathbb{E}\left[u(v, k) u(v, k)^\top \right] \delta = 0 \implies \delta = 0 \quad \text{if } \mathbb{E}\left[u(v, k) u(v, k)^\top \right] \text{ is invertible.}$$

3 Estimators for Common Functional Forms

In estimation, researchers can use standard proxy estimators with the additional constraint that the production function has constant returns to scale. For simple production functions, it is straightforward to derive closed form conditions. For more complex production functions, researchers can equivalently add the constant returns to scale restriction with additional moment restrictions in the generalized method of moments problem. Linear-in-parameters functional forms such as Cobb-Douglas and translog have particularly straightforward estimators. In this section, we offer a guide on how to apply our method to Cobb-Douglas and translog production functions in the framework of Ackerberg et al. (2015) used in De Loecker and Warzynski (2012). Appendix C presents the method for general linear-in-parameters production functions.

In all cases, the goal is to recover the flexible input elasticity f_v to then apply the optimization condition in 5. Note that since the firm does not see ϵ_t before their choice of v_t, we must apply *expected* output to the markup equation, i.e. $\exp(q_t - \epsilon_t)$ instead of $\exp(q_t) = Q_t$.

3.1 Cobb-Douglas Production Functions

The Cobb-Douglas production with constant returns to scale can be written as,

$$q_t = \theta_v v_t + \theta_k k_t + a_t + \epsilon_t = (1 - \theta_k)v_t + \theta_k k_t + a_t + \epsilon_t \tag{26}$$

Estimation uses the standard proxy estimator in Ackerberg et al. (2015):

1. Impose the CRS parameter restrictions on the production function as above.
2. Regress q_t on a specified transformation of (v_t, k_t) to estimate $\phi(v_t, k_t) = \mathbb{E}[q_t|v_t, k_t]$:

$$q_t = f + a_t + \epsilon_t = \phi(v_t, k_t) + \epsilon_t \tag{27}$$

3. For a given guess of θ, write $a_t = \phi(v_t, k_t) - f(v_t, k_t)$.
4. Regress a_t on a specified transformation of a_{t-1} to estimate $g(a_{t-1}) = \mathbb{E}[a_t|a_{t-1}]$:

$$a_t = g(a_{t-1}) + \eta_t \tag{28}$$

5. Solve the moment condition for θ:

$$\frac{1}{n}\sum_{it} k_{it}\eta_{it}(\theta) = 0 \tag{29}$$

Instead of the first step, the researcher may begin with the second step, but add the additional moment restriction:

$$\frac{1}{n}\sum_{it} RTS_{it}(\theta) - 1 = 0$$

3.2 Translog Production Functions

For the translog production function, we have:

$$q_t = \theta_v v_t + \theta_k k_t + \theta_{vv} v_t^2 + \theta_{kk} k_t^2 + \theta_{vk} v_t k_t + a_t + \epsilon_t$$

Constant returns to scale implies:

$$f_v + f_k = \theta_v + 2\theta_{vv} v + \theta_{vk} k + \theta_k + 2\theta_{kk} k + \theta_{vk} v = 1$$

which gives the following parameter restrictions:

$$\begin{array}{rcl} \theta_k + \theta_v & = & 1 \\ 2\theta_{vv} + \theta_{vk} & = & 0 \\ \theta_{vk} + 2\theta_{kk} & = & 0 \end{array}$$

Hence the translog production function has three parameters after imposing CRS, and researchers can proceed following the exact same steps as above.

Alternatively, if the researcher may skip the first step by adding the additional moment restrictions:

$$\begin{array}{rcl} \frac{1}{n}\sum_{it} v_{it}(RTS_{it}(\theta) - 1) & = & 0 \\ \frac{1}{n}\sum_{it} k_{it}(RTS_{it}(\theta) - 1) & = & 0 \\ \frac{1}{n}\sum_{it} RTS_{it}(\theta) - 1 & = & 0 \end{array}$$

4 Sizing the Problem: US Public Firm Markups

We apply our approach to the Fundamental Annual Compustat file from Wharton Research Data Services. These data span from 1951 to 2017 and cover private sector firms with public equity or debt. Compustat contains firm-level balance sheet information on: sales; operating expenses (OPEX); cost of goods sold (COGS); selling, general, and administrative expenses (SGA); capital (via net and gross plants, property, and equipment (PPE)); and industry classification.

To select domestic firms, we use standard industry format observations in USD with Foreign Incorporation Codes (FIC) in the US. For data quality, we include only observations with positive assets, sales, OPEX, and gross PPE. To avoid picking up merger and acquisition distortions, we exclude observations in which acquisitions are larger than 5 percent of total assets. Just under 9 percent of the sample are missing information on sales, OPEX, gross PPE, or net PPE, within a given firm. We replace these missing observations with a linear interpolation of their neighboring values.

For industry classifications, Compustat includes only the current SIC code and historical SIC codes starting in 1987. We use historical SIC codes when available. We backfill the industry classification with the first historical SIC code, and replace any remaining missing observations with the current SIC code. To better balance the number of observations in each industry which will be useful for later estimation, we map these SIC codes to the Fama-French 49 industry groups. This mapping is standard in the finance and accounting literature, and roughly combines similar industries that have few observations, and separates industries that have many observations into subindustries. Finally, we exclude utilities (Fama-French 49 code 31) because they are heavily regulated on prices, and financials (Fama-French 49 codes 45 to 49) because their balance sheets are dramatically different from other firms.

To get a quantity measure of sales and flexible inputs, we deflate sales, OPEX, COGS, and SGA by the GDP deflator (NIPA Table 1.1.9 Line 1).

As is standard in the production function estimation literature, we construct our measure of capital using the perpetual inventory method. Specifically, we initialize the capital stock using the first available entry of gross PPE. We then iterate forward on capital using the accumulation equation $K_{it} = K_{it-1} + I_{it-1} - \delta K_{it-1}$, where we compute net investment $I_{it-1} - \delta K_{it-1}$ using changes to net PPE. Since we want a quantity measure of the capital stock, we deflate net investment by the nonresidential fixed investment deflator (NIPA Table 1.1.9 Line 9).

Our goal in this paper is not to determine the appropriate specification for production technology using Compustat data. For more information on this debate, see Traina (2018) and Diez et al. (2018). Rather, we believe it's important to evaluate the impact of nonidentification on a variety of applied specifications. We will focus on the simplest specification (Cobb-Douglas technology with flexible OPEX), but also present results for three others when relevant. In total, we vary {Cobb-Douglas; translog} x {OPEX and PPE; COGS, SGA, and PPE}. Our estimator typically increasingly outperforms as we increase the complexity of the specification.

Denote $x = \log OPEX$, $c = \log COGS$, $s = \log SGA$, and $k = \log PPE$. Our specifications for the production function are:

$$\text{CD-OPEX} : \theta_x x + \theta_k k \tag{30}$$

$$\text{CD-COGS} : \theta_c c + \theta_s s + \theta_k k \tag{31}$$

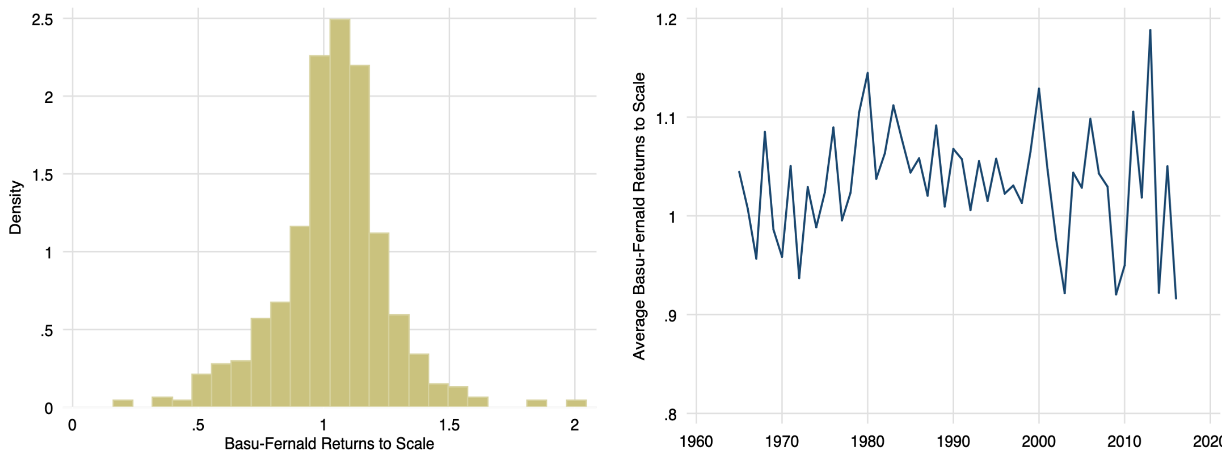
$$\text{TL-OPEX} : \theta_x x + \theta_k k + \theta_{xx} x^2 + \theta_{kk} k^2 + \theta_{xk} xk \tag{32}$$

$$\text{TL-COGS} : \theta_c c + \theta_s s + \theta_k k + \theta_{cc} c^2 + \theta_{ss} s^2 + \theta_{kk} k^2 + \theta_{cs} cs + \theta_{sk} sk + \theta_{ck} ck \tag{33}$$

Since we will proceed by imposing constant returns to scale with our benchmark estimator, we first check whether this assumption is reasonable in this particular data setting. To do so, we follow Basu and Fernald (1997), which recommends the following procedure:

1. Generate the cost share of total costs for each input.
2. Generate a composite input growth index that sums each input's growth by its lagged cost share.
3. Regress output growth on this composite input growth; the coefficient is an estimate of the returns to scale.

Figure 1: Basu-Fernald Returns to Scale Estimates



Source: Data from Compustat, estimates from authors' calculations.

Although this algorithm does not identify the true returns to scale under typical conditions, it does offer an approximation that has been used in practice (e.g. Syverson (2004)). We follow the procedure for each industry in each year to produce a distribution of returns to scale estimates. Figure 2 collects the results. The left panel presents the distribution of scale elasticity estimates, and the right panel shows the time series of their cross-sectional averages.

The left panel shows that the typical industry-year exhibits constant or slightly increasing returns to scale, with a mean of 1.04. There is some variation away from our CRS benchmark, with a standard deviation of 0.22, which likely represents a combination of true differences in scale elasticities, sampling error, and misspecification. The right panel shows that there is no discernible time trend in these estimates. Regressing our scale elasticity estimates on a linear time trend returns a coefficient that is statistically indistinguishable from zero.

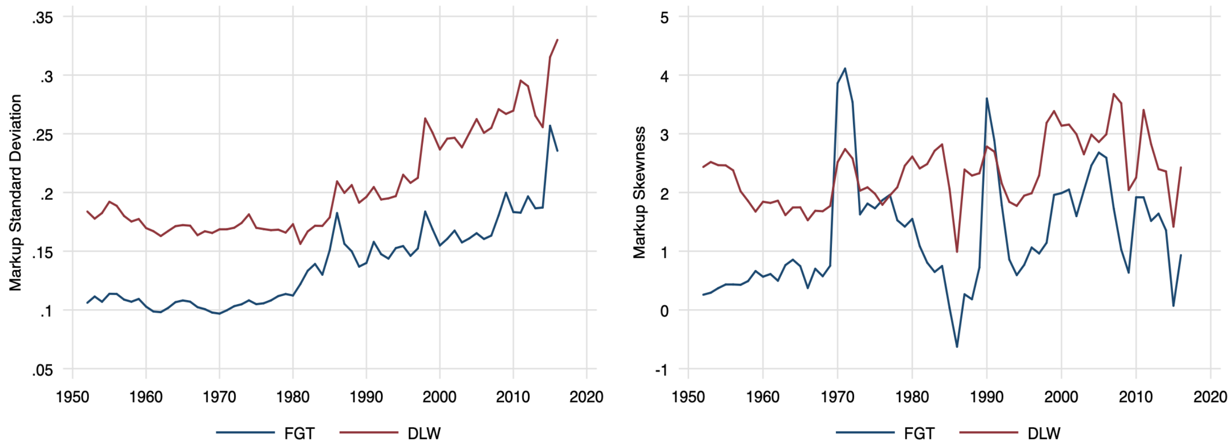
To level the playing field away from obvious outliers, the rest of the paper subsets on markups between 0.1 and 10, which we consider to be a priori reasonable bounds for anomalies.

4.1 Heterogeneous Markups

To give the DLW estimator its best shot, we move forward with the simplest specification (CD-OPEX). Figure 6 shows the differences in markup distribution by estimator. The left panel measures the cross-sectional standard deviation, whereas the right panel measures the cross-sectional skewness (both weighted by OPEX). In the left panel, the blue line is stable at about 0.10 from 1951 to 1970, when it rises steadily to about 0.20 in 2010. This distribution occurs around a mean of about 1.15 throughout the sample. The rise is notably linear, and there are no otherwise obvious patterns beforehand.

The red line in the left panel is similar in the qualitative pattern, yet at a different y-axis scale. In this case, the blue line starts relatively flat between 0.15 and 0.20 from 1951 to 1980, when it also rises to about 0.30 in 2017. Although the fact that it captures

Figure 2: Variation in Markups by Estimator



Source: Data from Compustat, estimates from authors' calculations.

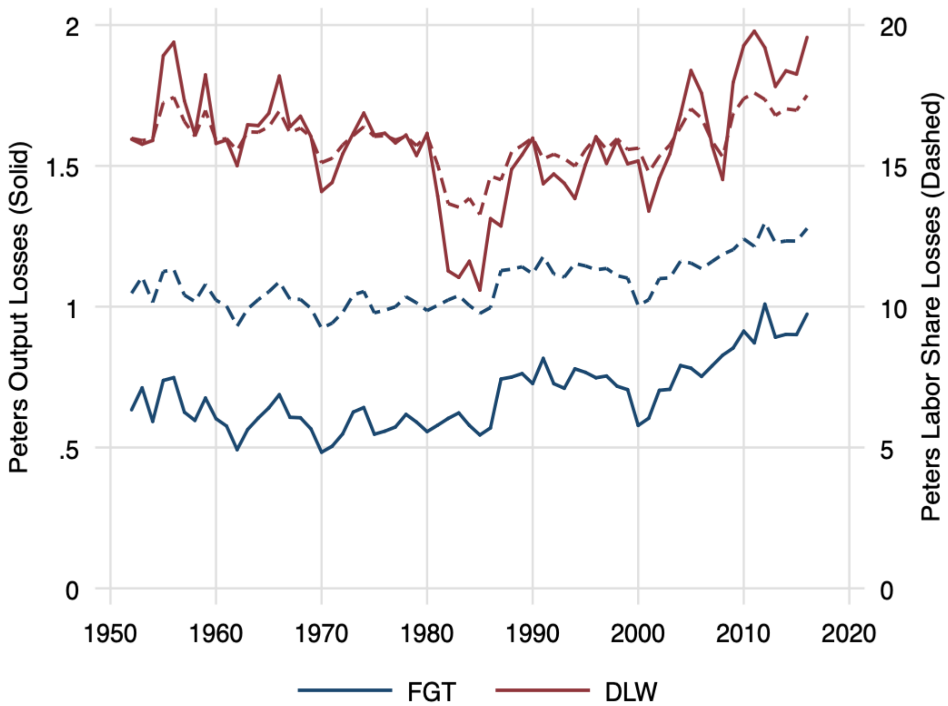
a comparable pattern is reassuring that we're measuring important underlying economic phenomena, the quantification deviates from our identified model in two important ways. First, it meaningfully overstates the level of dispersion, particularly in the earlier parts of the sample by about 50 percent. Second, because of this overstatement, it significantly underestimates the relative size of the increase by about 50 percent. The right panel shows the cross-sectional skewness of the two estimators. Our CRS estimator is a bit more volatile year-to-year, but typically at a lower level than the DLW estimator. The two estimators do not exhibit any time series trend.

4.2 How Costly is Nonidentification?

To benchmark how significant these results are, we use the sufficient statistic approach described in Peters (2018) to estimate the corresponding implied output losses from misallocation. The model builds on the canonical Klette and Kortum (2004) framework by adding imperfect competition in product markets to generate variable markups. Given a Pareto distribution of markups, we can approximate the losses to output and the labor share based on the Pareto parameter. Hence to match our estimates to the model, we fit a Pareto distribution of markups each year, and use the model to back out the implied losses.

Figure 7 reports our results. Here we see the stark differences in welfare evaluation and macroeconomic implications based on nonidentification. Our identified estimator shows output losses hovering just above 0.5 percent annually, then rising to 1 percent today. For reference, 1 percent of US output today is about $200B, about half the budget of the US Department of Defense. While this statistic is large, the DLW estimator implies output losses of roughly twice the scale. The labor share comparison is less stark; we estimate a labor share that's about 11 percent lower than an idealized benchmark because of markups, whereas the DLW approach estimates about 17 percent.

Figure 3: The Cost of Markups



Source: Data from Compustat, estimates from authors' calculations using the sufficient statistics of Peters (2018).

Table 1: Specification Error using Bounds, Raw Percent

Specification		Too Low		Too High	
Form	Flex	FGT	DLW	FGT	DLW
CD	OPEX	0.16	0.04	0.03	0.55
CD	COGS	0.31	0.35	0.41	0.42
TL	OPEX	0.18	0.51	0.62	0.07
TL	COGS	0.41	0.00	0.00	0.45

Table 2: Specification Error using Bounds, Average Violation

Specification		Too Low		Too High	
Form	Flex	FGT	DLW	FGT	DLW
CD	OPEX	0.02	0.00	0.00	0.06
CD	COGS	0.04	0.06	0.11	0.02
TL	OPEX	0.02	4.99	0.06	0.23
TL	COGS	0.13	0.00	0.00	0.27

4.3 Specification Testing with Partial Identification Bounds

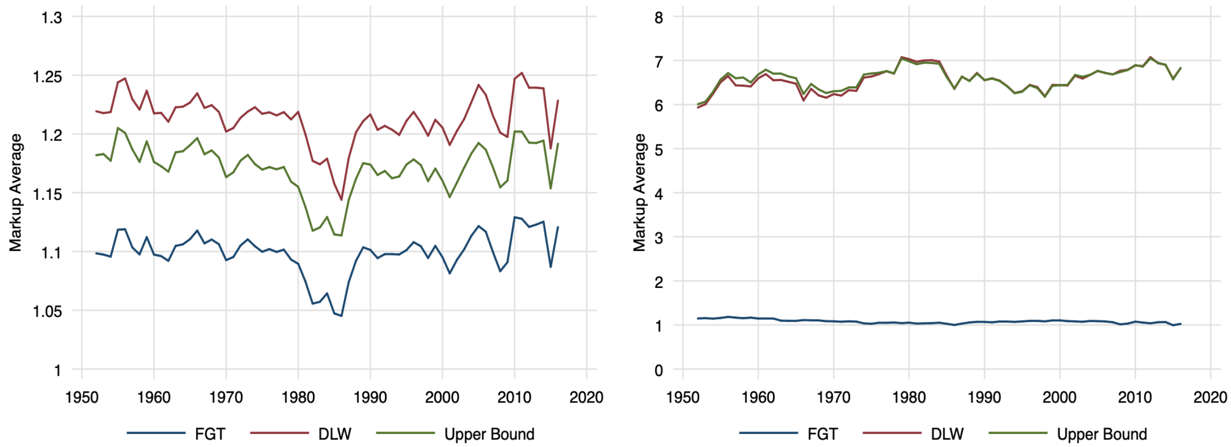
Table 1 summarizes how the estimators perform relative to the partial identification bounds introduced in Section 1.2. Each row represents a different production function specification: {Cobb-Douglas; translog} x {OPEX and PPE; COGS, SGA, and PPE}. The "Too Low" columns report the share of markups below the lower bound of 1 under each estimator, whereas the "Too High" columns report the share above the upper bound. We aggregate these shares on a cost-weighted basis, reasoning that the flexible input cost is the relevant weight for welfare calculations per Edmond et al. (2018).

Our estimator typically performs substantially better at not exceeding the upper bound, and performs slightly worse at not falling below the lower bound. However, this takeaway varies across production function specification. While Table 3 shows how often markups fall outside the partial identification bounds, it does not show how substantial these violations are. To check the quantitative importance of these bound violations, we next look at the average bias they induce. We measure each markup estimate's distance past the bound, setting within-bound estimates to zero. Table 4 presents the averages of these bound violations.

While our estimator still has some notable biases, a key feature is that they are not substantial. The largest average bias produced is 0.13, and a typical bias is about 0.05. In contrast, the DLW estimator produces a dispersed set of biases, including some very large ones.

Tables 1 and 2 do not show the economic significance of this evaluation in levels. Intuitively, an estimator that exceeds an upper bound of 1.5 is meaningfully different from an estimator that exceeds an upper bound of 5. For all but the TL-COGS specification, the bounds are reasonably tight (below 1.5 on average); for TL-COGS, however, the bounds exceed 5. By comparison, the seminal paper Chaloupka (1991) reports a demand elasticity for cigarette

Figure 4: The Level of Markups by Estimator



Source: Data from Compustat, estimates from authors' calculations.

smokers of about -0.5, which implies a monopolist markup of 2.

4.4 Stability Across Specifications

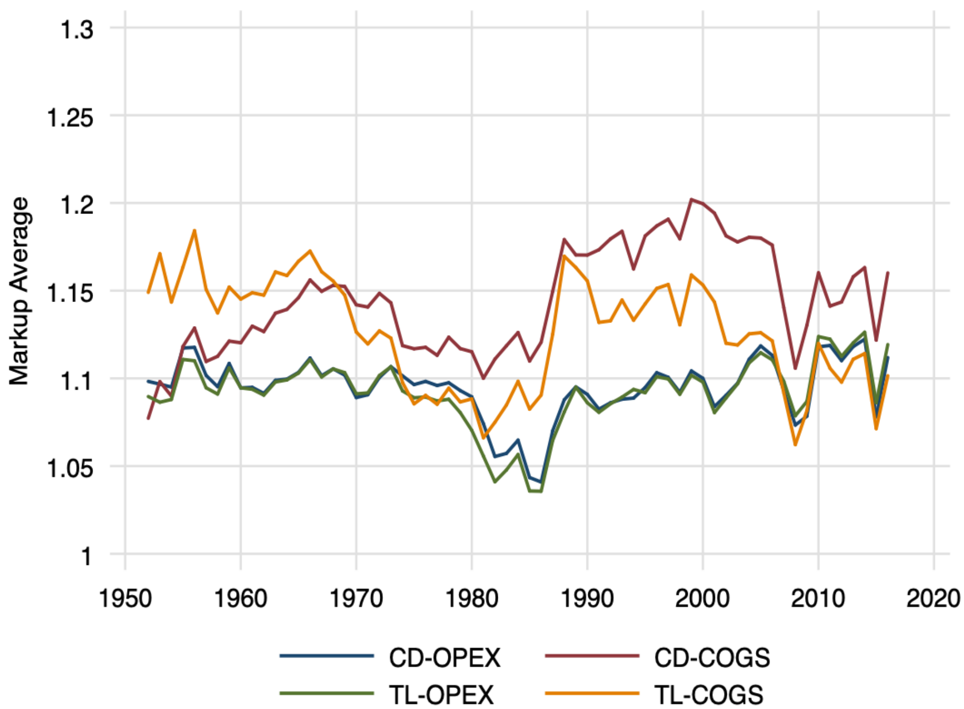
Motivated by the partial identification results, we next explore the stability of the estimators by looking at how their resultant markups change across the different production technology specifications. Figure 3 presents the time series of aggregate markup estimates for the four considered in this paper. The left panel is the simplest specification – Cobb-Douglas OPEX as used in Traina (2018). The right panel is the most complex specification – translog COGS-SGA, similar to the robustness section of Diez et al. (2018). The blue line represents our CRS estimator, the red line represents the DLW estimator, and the green line represents the upper bound implied by Flynn (2019) and Flynn and Gandhi (2018). As before, these estimates use flexible input cost-weights to aggregate.

As hinted by the earlier tables, the DLW estimator is typically close to or exceeds the theoretical upper bound. In the left panel, the red line is about 0.05 higher than the green line, showing that the DLW estimator's upward bias completely swamps any downward bias. More notably for these figures, however, is just how high the theoretical upper bound can reach. In the right panel, both the red and green lines fluctuate between 6 and 7, much higher than in the left panel and more generally implausibly high. Our CRS estimator, by comparison, looks about the same across specifications.

Figure 4 confirms this stability by collecting our results across the four specifications. Our estimator is remarkably stable to different specifications of the production function. The differences in aggregate means are low, typically within 0.05 to 0.10.

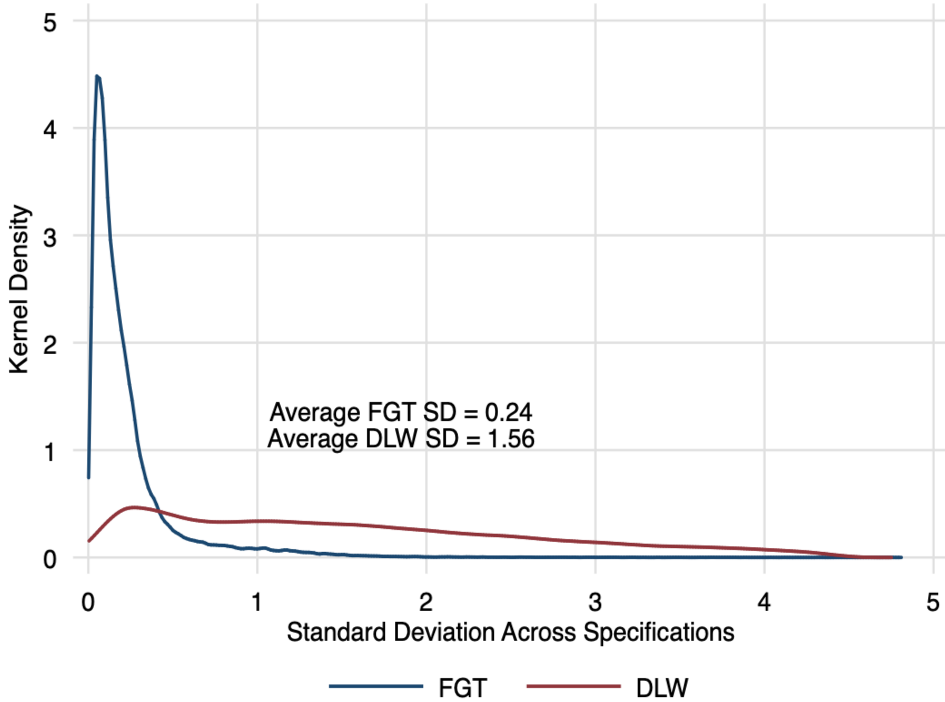
These estimates are at the aggregate level, however, which is more useful for stylized facts but less useful for firm-level analyses. To quantify how much the markup estimates vary at the micro level, we calculate the standard deviation across the four specifications for each estimator. We therefore have a dispersion estimate for each firm-year observation. Figure 5

Figure 5: The Stability of the FGT Estimator



Source: Data from Compustat, estimates from authors' calculations.

Figure 6: How Much Do Firm-Level Markups Vary by Specification?



Source: Data from Compustat, estimates from authors' calculations.

presents the distribution of these dispersion estimates for the two estimators.

Our estimator typically produces markups that are robust to alternative production technology specifications. The average standard deviation is 0.24, against 1.56 for the DLW estimator. And recall that these statistics are after truncating the distributions at 0.1 and 10 to remove anomalous outliers, which plague the DLW but not our CRS estimator.

5 Sizing the Problem: A Calibrated Macro Monte Carlo

In this section, we use a Monte Carlo experiment to demonstrate what can go wrong if we ignore the identification problem and measure markups using lagged flexible inputs as instruments. By comparing our estimator to the De Loecker and Warzynski (2012) estimator, we show that inference on the pattern of markups in an industry can be misleading if we do not deal with the identification problem. We also show that reasonable deviations from constant returns to scale (where the true returns to scale is not 1 but we assume it is) have less effect on inference than ignoring the identification problem entirely. If we do not believe that the returns to scale are far from 1, these results suggest that small deviations from constant returns to scale will not overly affect inference.

We construct a data generating process that is tractable to compute and resembles standard macro models and parameter choices. We compute the objects of interest in our actual application (aggregate cost-weighted markups) and study how well our estimator compares to the status quo DLW estimator used in the literature. In our data generating process, there are several industries each of which faces its own constant elasticity demand curve (the representative consumer has quasi-linear preferences). Each industry is served by a monopolist firm. We observe many markets like this giving us a number of firms to use in each industry to estimate the production function. There is a markup distribution because each industry has a different constant elasticity demand curve.

Within each industry, a monopolist solves the following dynamic programming problem,

$$M(A,K) = \max_{V,X,Q_E} Q_E^{\eta+1}\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right] - W_X X - W_V V + \beta\mathbb{E}\left[M\left(A^{0.9}\nu, \delta\times K + X\right)\right] \tag{34}$$

$$\text{ST:}\quad A\times V^{\theta}K^{\gamma} \geq Q_E, \tag{35}$$

where the expectation is over ν, a shock to productivity. The dynamic programming problem can be solved analytically, so it is straightforward to draw data from the data generating process.

We parametrize our experiment at the following parameter values (for a given θ , γ, and η),[9]

$$Q = Q_E\times\exp(\epsilon), \quad Q = V^{\theta}K^{\gamma}A, \quad \log\nu\sim N(0,0.25) \tag{36}$$

$$\beta = 0.96, \quad \delta = 0.9, \quad W_X = 1, \quad W_V = 1 \tag{37}$$

$$\epsilon \sim N(0,0.5). \tag{38}$$

At these parameter values, the true markup within an industry is always $-\eta^{-1}$. We simulate firm decisions within each of 20 industries in a population of 200 markets for 10 time periods (each market has a different firm being the monopolist for that industry). Each firm's initial capital and productivity comes from the following distribution:

$$\log(K_0 - 1)\sim N(0,1) \tag{39}$$

$$\log A_0 \sim N(0,1). \tag{40}$$

We estimate the model using data from $t = 3$ to $t = 10$ (because the DLW estimator will need two lags of flexible inputs).

Our goal is to compare our CRS estimator against the DLW estimator, which is currently the literature's standard for markup estimation. We study how well each method does at estimating average markups and the standard deviation of markups. Table 3 reports the bias of each method across 10000 Monte Carlo simulations for different values of (θ,γ), some of which satisfy $\theta+\gamma = 1$ and some of which do not.

To be precise, we use the same proxy assumptions for both the DLW estimator and our proposed estimator based on constant returns to scale,

$$q_{it} = \theta v_{it} + \gamma k_{it} + a_{it} + \epsilon_{it} \tag{41}$$

$$a_{it} = \rho a_{it-1} + \log\nu_{it}. \tag{42}$$

[9]For full details on the solution to the dynamic programming problem, see Appendix D.

Table 3: Monte Carlo Results

		Percent Bias in Markup Statistics			
		Mean		Std. Dev.	
θ	γ	FGT	DLW	FGT	DLW
0.865	0.135	0.44%	8.97%	0.35%	5.65%
[0.7785, 0.9515]	[0.1215, 0.1485]	2.08%	9.79%	0.53%	5.88%

FGT refers to the method proposed in this paper. DLW refers to applying the proxy method using lagged flexible inputs to instrument for current flexible inputs. Elasticity intervals are random variables drawn from a uniform distribution for each industry.

What is different between the two models are the moments used in estimation. The identifying moments for both models are,

$$\textbf{DLW Moments:} \quad \mathbb{E}\left[\begin{pmatrix} k_{it} \\ v_{it-2} \end{pmatrix} \log \nu_{it}\right] = 0 \tag{43}$$

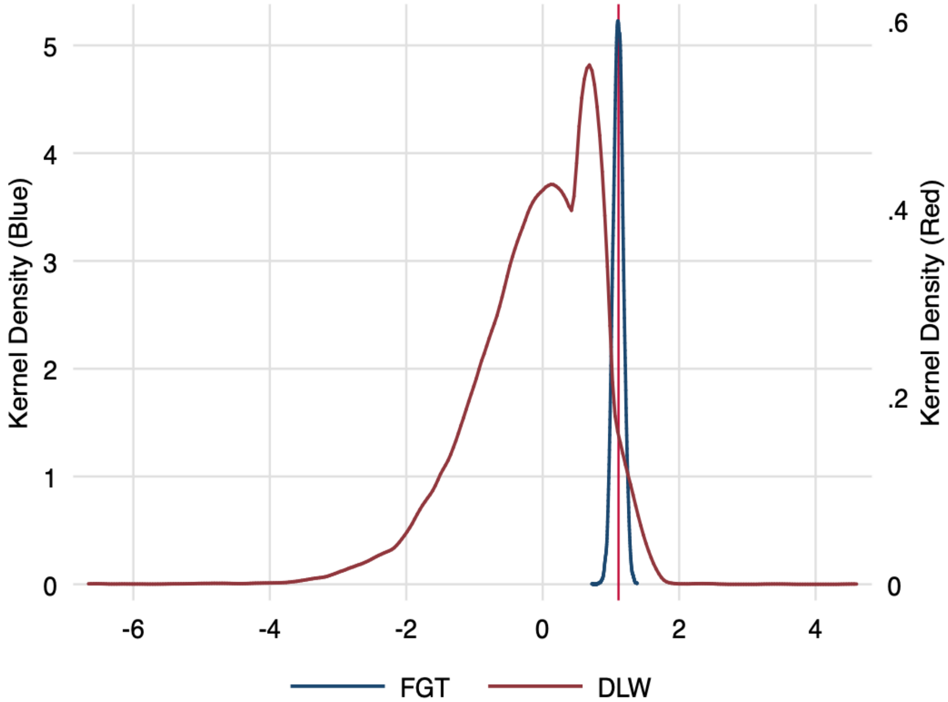
$$\textbf{FGT Moments:} \quad \mathbb{E}\left[k_{it} \log \nu_{it}\right] = 1 - \theta - \gamma = 0 \tag{44}$$

We draw the demand elasticity parameter for each industry from the uniform distribution on $[-0.5, -0.2]$. We use the same parameters for each of the twenty industries in all Monte Carlo experiments.

Table 3 collects the results of our Monte Carlo experiment. Each row represents a different parametrization of the flexible input elasticity θ and the capital elasticity γ. In the first row, our estimator is the correct specification, constant returns to scale, with the flexible output elasticity chosen to match Edmond et al. (2018). In the second row, we add some noise to the returns to scale so that some industries has increasing returns to scale and others have decreasing returns to the scale. The second row demonstrates that our results do not hinge on returns to scale being exactly 1. The first set of columns presents the average bias in the level of markups (weighted by variable input cost) as a percentage of the true value; the second set of columns presents the average bias in the dispersion of markups (weighted by variable input cost) also as a percentage of the true value. We measure bias relative to the estimator we would use if we knew the true production function (to avoid measuring as bias the fact that we only observe ex-post output instead of ex-ante output) — that is, we apply the markup formula replacing the flexible input elasticity with the true flexible input elasticity.

We find the bias from ignoring the identification problem is about five as large as the bias from being slightly wrong about the returns to scale in the industry under misspecification of the returns to scale. The bias of the DLW cost-weighted markup estimate is 9.0 percent under CRS and 9.8 percent under CRS-plus-noise.. In contrast, the bias from our CRS estimator is small, ranging from 0.44 percent under correct specification to 2.08 percent under misspecification. Our estimate of the standard deviation of markups is also slightly better — even under misspecification (when the RTS is not one) — and recall: the standard

Figure 7: Markup Distribution by Estimator



Source: Data from a specified generating process closely following Edmond et al. (2018), estimates from authors' calculations. The vertical line marks the true average.

deviations are around two different means so the distribution of markups is much better captured using our approach.

Figure 7 shows the full distribution of cost-weighted average markup estimates for both the DLW and our CRS estimators under constant returns to scale for a single industry. This graph shows visually that the DLW estimator is not identified because the distribution is bi-modal. In contrast, our CRS estimator is normally-distributed around the true value.

5.1 Input Price Variation Is Not a (Practical) Solution

De Loecker and Warzynski (2012) (and more recent work by De Loecker and Eeckhout 2017 and Traina (2018)) uses lags of flexible inputs to instrument for current flexible inputs. That paper's argument is that lags of flexible inputs are correlated with the price of the firm's inputs which will shift the distribution of current inputs. And, since lagged inputs were chosen in the past, they are uncorrelated with later shocks to productivity. Unfortunately, the use of lagged inputs and appeal to unobserved input price variation does not overcome the GNR identification problem.

The basic reason instrumental variable methods based on unobserved input price variation (or other unobserved firm heterogeneity) fail is that either:

1. There is input price variation so we have an omitted variable (input price) in the residual because input price should be included in the proxy function. So our instrument is correlated with the residual (our instrument is invalid).
2. There is no input-price variation so lagged flexible inputs have no power (our instrument is weak).

In the following section, we establish this point.

If there is no input price variation, then Gandhi et al. (2017) have already shown that the model is not identified as we review in Section 1. The problem is fundamental: for lagged inputs to satisfy the exclusion restriction, they must be uncorrelated with η_{it}, but for lagged inputs to have strength as instruments, they must be correlated with η_{it}.

Now consider the case where input prices vary by firm, yielding another firm-specific state variable aside from η_{it} that might shift v_{it}. There remains two general problems: (1) the nature of the input price variation required is more demanding than we might intuit (it is not enough that input prices vary by firm); and (2) the nature of the data required is more demanding than we might see in practice (we must observe the firm-level input prices).

First, suppose input prices vary across firms, and we observe these input prices. In that case, flexible input demand is,

$$v_t = v_t\left(a_t, k_t, w_t\right) = v_t\left(g\left(v_{t-1}, k_{t-1}, w_{t-1}\right) + \eta_t, k_t, w_t\right), \tag{45}$$

where w_t is the log price of flexible inputs.

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

To see why this is a stronger condition than we might think, suppose w_t is an AR(1) process,

$$w_{it} = \rho^w w_{it-1} + \text{Innovation}_{it}, \quad \text{Innovation}_{it} \sim N\left(0, 1\right). \tag{46}$$

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

What we need for v_{it-2} to be an instrument is for there to be a firm-specific component of the wage process. For example,

$$w_{it} = \rho^w w_{it-1} + \xi_i + \text{innovation}_{it}. \tag{47}$$

Or that w_t is at least an AR(2) process; we need an underlying state variable that affects w_{it} aside from w_{it-1} for which v_{it-2} can proxy.

The second question for identification is whether v_{t-2} is a valid instrument, i.e. whether its variation is actually exogenous. This question might seem irrelevant because we have already assumed v_{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 related to innovations in the productivity process—when wages go up, it is because the firm is using more productive inputs—in which case v_{t-2} is correlated with η_t.

But the larger issue is that we often do not observe w_t. If input prices vary across firms but we do not observe this variation, then using twice-lagged[10] flexible inputs as instruments will not work. In this case, the proxy function is:

$$v_t = v_t(a_t, k_t, w_t) \implies a_t = v_t^{-1}(v_t, k_t, w_t) \tag{48}$$

$$\implies q_t = f(v_t, k_t) + g(v_{t-1}, k_{t-1}, w_{t-1}) + \eta_t + \epsilon_t. \tag{49}$$

If we do not observe w_{t-1} and we estimate 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 is typically hard to observe for intermediate inputs. In fact, the reason to use lagged flexible inputs as the instrument is that we do not observe firm wages—otherwise, if we believe wage variation is excluded, we could just use firm-level wages (w_{it}) directly as the instrument and it would likely be a stronger instrument than indirectly proxying w_{it} via v_{it-2}.

We also compare the two estimators when we introduce firm-level input price variation to the data-generating process. Both estimators are misspecified under this data generating process, but the DLW instruments are no longer weak because lagged inputs are correlated with input price variation; the bias shifts from being a result of the instruments having no power to an omitted variable bias. The data generating process we use for this Monte Carlo experiment is the same as above except that,

$$W_V \sim Uniform(1-\kappa, 1+\kappa),$$

where, previously, we had set $W_V = 1$. In Table 4 we present these results. The structure of the table is the same as before, but now with input price variation.

[10] 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, first-lagged materials are not excluded, so we need to use twice-lagged materials (flexible inputs).

Table 4: Monte Carlo Results with Input Price Variation

			Percent Bias in Markup Statistics			
			Mean		Std. Dev.	
κ	θ	γ	FGT	DLW	FGT	DLW
0.05	0.865	0.135	0.42	8.89	0.31	5.49
	[0.7785, 0.9515]	[0.1215, 0.1485]	2.03	9.67	0.44	5.77
0.10	0.865	0.135	0.28	9.29	0.18	5.93
	[0.7785, 0.9515]	[0.1215, 0.1485]	1.87	9.77	0.33	5.98
0.50	0.865	0.135	4.58	11.98	4.34	9.11
	[0.7785, 0.9515]	[0.1215, 0.1485]	2.89	12.68	4.77	9.58

Even under this misspecification which gives lagged flexible inputs power in the first stage (the second stage is still misspecified), our estimator performs better. Intuitively, this is because as the variance of input prices becomes small, our estimator is a better and better approximation to the truth while the DLW estimator is not because it is not identified. There is a set of assumptions under which our estimator is identified and it appears fairly robust to small deviations from those assumptions.

6 Evaluating Estimation Extensions

6.1 Variable Returns to Scale

We can extend our identification results for at least two common classes of production functions where returns to scale are not identically one but are a function of fixed input use (inputs that are uncorrelated with productivity innovation shocks). We make the following two assumptions, relaxing constant returns to scale:

Assumption 2. Returns to scale is a function $h(k)$, that depends only on fixed inputs.

$$f_v(v,k) + f_k(v,k) = h(k). \tag{50}$$

Assumption 3. The *average* return to scale is 1 (or some other known number),

$$\mathbb{E}\left[h(k_{it})\right] = 1 \tag{51}$$

6.1.1 The Separable Case

Suppose that $f = f^1(v) + f^2(k)$, a log additively-separable production function, and suppose that the returns to scale of this production function is a function, $h(k)$. Then,

$$f^1_v(v) + f^2_k(k) = h(k). \tag{52}$$

Differentiating the above identity with respect to v gives,

$$f^1_{vv}(v) = 0. \tag{53}$$

So these assumptions together imply that the production function has a constant elasticity in v, so we have the "partially linear production function",

$$f = \theta_v v + f^2(k). \tag{54}$$

This production function is a generalization of the Cobb Douglas production function.

By Assumption 3, we have,

$$\mathbb{E}\left[\theta_v + f_k^2(k_t)\right] = 1 \tag{55}$$

$$\implies \theta_v = 1 - \mathbb{E}\left[f_k^2(k_t)\right] \tag{56}$$

From the proxy structure, using the same algebra as in our proof for the CRTS result, we know that we can write f_k as a function of f_v,

$$f_k = \frac{\partial}{\partial k}\mathbb{E}\left[q_t - \int_{\overline{v}}^{v_t} f_v(v', k_t)\, dv' | k_t = k, v_{t-1}, k_{t-1}\right] + \int_{\overline{v}}^{v} f_{vk}(v', k)\, dv'. \tag{57}$$

Plugging in the partially linear production function specification to this equation gives,

$$f_k^2 = \frac{\partial}{\partial k}\mathbb{E}\left[q_t - \theta_v(v_t - \overline{v})\,|k_t = k, v_{t-1}, k_{t-1}\right] \tag{58}$$

So, we have,

$$\theta_v = 1 - \mathbb{E}\left[f_k^2(k_t)\right] \tag{59}$$

$$= 1 - \mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[q_t|k_t = k_t, v_{t-1}, k_{t-1}\right]\right] + \theta_v \mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[v_t|k_t = k_t, v_{t-1}, k_{t-1}\right]\right]. \tag{60}$$

We need one (testable) assumption to ensure identification. The assumption rules out that $v_t \approx k_t + \text{noise}$.

Assumption 4. $\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[v_t|k_t = k_t, v_{t-1}, k_{t-1}\right]\right] \neq 1$

Then,

$$\theta_v = \frac{1 - \mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[q_t|k_t = k_t, v_{t-1}, k_{t-1}\right]\right]}{1 - \mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[v_t|k_t = k_t, v_{t-1}, k_{t-1}\right]\right]} \tag{61}$$

So θ_v is identified. $f_k^2(k)$ is then identified by,

$$f_k^2 = \frac{\partial}{\partial k}\mathbb{E}\left[q_t - \theta_v(v_t - \overline{v})\,|k_t = k, v_{t-1}, k_{t-1}\right], \tag{62}$$

and the entire production function is identified.

6.1.2 The Translog Case

If we want to allow the output elasticity of the inputs to vary with the use of the other input, we can use a translog specification. The translog production function will also be identified by Assumptions 2 and 3, but with a different rank condition than used for the partially linear model.

The translog production function is,

$$f = \theta_v v + \theta_{vk} vk + \frac{1}{2}\theta_{vv} v^2 + \theta_k k + \frac{1}{2}\theta_{kk} k^2. \tag{63}$$

So returns to scale are,

$$\theta_v + \theta_{vk} k + \theta_{vv} v + \theta_k + \theta_{vk} v + \theta_{kk} k. \tag{64}$$

From 2, we have assumed the returns to scale function does not vary with v so the coefficient on v above must be zero, or,

$$\theta_{vv} = -\theta_{vk}. \tag{65}$$

This restriction is also required for translog production functions to have constant returns to scale.

Applying the assumption that the returns to scale is 1 on average gives,

$$\theta_v + \theta_{vk}\mathbb{E}k - \theta_{vk}\mathbb{E}v + \theta_k + \theta_{vk}\mathbb{E}v + \theta_{kk}\mathbb{E}k = 1 \tag{66}$$

$$\implies \theta_v = 1 - \theta_{vk}\mathbb{E}k - \theta_k - \theta_{kk}\mathbb{E}k. \tag{67}$$

So we can write the production relationship, using the proxy structure as,

$$q_t = [1 - \theta_{vk}\mathbb{E}k_t - \theta_k - \theta_{kk}\mathbb{E}k_t]\, v_t + \theta_{vk} v_t k_t - \frac{1}{2}\theta_{vk} v_t^2 + \theta_k k_t + \frac{1}{2}\theta_{kk} k_t^2 \tag{68}$$

$$+\tilde{g}\left(v_{t-1}, k_{t-1}\right) + \eta_t + \epsilon_t \tag{69}$$

$$\implies q_t - v_t = \theta_{vk} v_t \left[k_t - \mathbb{E}k_t - \frac{1}{2}v_t\right] + \theta_k\left(k_t - v_t\right) + \theta_{kk}\left[\frac{1}{2}k_t^2 - v_t\mathbb{E}k_t\right] \tag{70}$$

$$+\tilde{g}\left(v_{t-1}, k_{t-1}\right) + \eta_t + \epsilon_t \tag{71}$$

The basic idea behind identification is that k_t affects each of the three terms with coefficients $(\theta_{vk}, \theta_k, \theta_{kk})$ in a linearly independent way. Take expectations and a derivative with respect to k_t,

$$\frac{\partial}{\partial k_t}\mathbb{E}\left[q_t - v_t | k_t, k_{t-1}, v_{t-1}\right] = \theta_{vk}\left\{\frac{\partial}{\partial k_t}\mathbb{E}\left[v_t\left(k_t - \mathbb{E}k_t\right)|\cdot\right] - \frac{1}{2}\frac{\partial}{\partial k_t}\mathbb{E}\left[v_t^2|\cdot\right]\right\} \tag{72}$$

$$+\theta_k\left(1 - \frac{\partial}{\partial k_t}\mathbb{E}\left[v_t|\cdot\right]\right) + \theta_{kk}\left[k_t - \mathbb{E}k_t \times \frac{\partial}{\partial k_t}\mathbb{E}\left[v_t|\cdot\right]\right]. \tag{73}$$

Write the coefficient weighting θ_{vk} as $\tau_{vk}(k_t, v_{t-1}, k_{t-1})$, the coefficient weighting θ_k as $\tau_k(k_t, v_{t-1}, k_{t-1})$, and the coefficient weighting θ_{kk} as $\tau_{kk}(k_t, v_{t-1}, k_{t-1})$ so that,

$$\frac{\partial}{\partial k_t}\mathbb{E}[q_t - v_t | k_t, k_{t-1}, v_{t-1}] = \tag{74}$$

$$\theta_{vk}\tau_{vk}(k_t, k_{t-1}, v_{t-1}) + \theta_k\tau_k(k_t, k_{t-1}, v_{t-1}) + \theta_{kk}\tau_{kk}(k_t, k_{t-1}, v_{t-1}) \tag{75}$$

We then have the following rank condition to identify the translog production function,

Assumption 5.

$$\mathbb{E}\left[(\tau_{vk}, \tau_k, \tau_{kk})(\tau_{vk}, \tau_k, \tau_{kk})^{\top}\right] \text{ is invertible.} \tag{76}$$

Assumption 5 is similar to the identification condition for CRTS and for the partially linear production function. It essentially rules out $\mathbb{E}[v_t|k_t] = k_t$ and $\mathbb{E}[v_t^2|k_t] = k_t^2$. If Assumption 5 is true than the coefficients can be estimated by the regression,

$$\frac{\partial}{\partial k_t}\mathbb{E}[q_t - v_t | k_t, k_{t-1}, v_{t-1}] = \theta_{vk}\tau_{vk} + \theta_k\tau_k + \theta_{kk}\tau_{kk} + \text{estimation error.} \tag{77}$$

7 Concluding Remarks

Recent work in macroeconomics and industrial organization uses the firm's first order condition on flexible inputs to estimate markups from production data. This method offers a direct measure of market power, unlike market concentration or profitability which conflates market power and productivity. Moreover, it does not impose assumptions on demand or market structure, and is therefore consistent with a broad class of economic models.

However, this approach falls victim to the nonidentification critique in Gandhi, Navarro, and Rivers (2017) – the parameters in the first order condition are not identified by the most popular production estimators. Worse still, existing solutions to the nonidentification problem do not carry over to the setting of market power and imperfect competition. In practice, this problem means that statistical software will return parameters that use variation from pure specification error, making the results noninterpretable.

In this paper, we present a solution that relies on specifying the returns to scale of production. Under such an assumption, we solve the identification problem, and consequently offer applied researchers a practical tool to reliably measure market power in production datasets. We recommend benchmarking results with constant returns to scale, which prior work has found to be a good approximation of reality (Basu and Fernald (1997), Syverson (2004), Foster et al. (2008)). Researchers can then check the robustness of their findings through sensitivity analysis.

We close the paper by showing that remaining problems in production estimators may improve with emerging econometric methods that account for heterogeneous technology, offering a new direction for future research. We present a start in this direction with classification methods in Appendix E.

References

Ackerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica 83*(6), 2411–2451. 2, 3, 3.1

Asker, J., A. Collard-Wexler, and J. De Loecker (2014). Dynamic inputs and resource (mis)allocation. *Journal of Political Economy 122*(5), 1013–1063. (document)

Atkeson, A. and A. Burstein (2008). Pricing-to-market, trade costs, and international relative prices. *American Economic Review 98*(5), 1998–2031. (document)

Barkai, S. (2017). Declining labor and capital shares. Technical report, Stigler Center for the Study of the Economy and the State New Working Paper Series. (document)

Basu, S. and J. G. Fernald (1997). Returns to scale in US production: Estimates and implications. *Journal of Political Economy 105*(2), 249–283. (document), 4, 7

Bloom, N., R. Sadun, and J. Van Reenen (2012). Americans do IT better: US multinationals and the productivity miracle. *American Economic Review 102*(1), 167–201. (document)

Chaloupka, F. (1991). Rational addictive behavior and cigarette smoking. *Journal of Political Economy 99*(4), 722–742. (document), 4.3

De Loecker, J. and J. Eeckhout (2017). The rise of market power and the macroeconomic implications. 3, 5.1

De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review 102*(6), 2437–71. (document), 3, 5, 5.1, 5.1, 10

Diez, F., D. Leigh, and S. Tambunlertchai (2018). Global market power and its macroeconomic implications. 4, 4.4

Edmond, C., V. Midrigan, and D. Y. Xu (2015). Competition, markups, and the gains from international trade. *American Economic Review 105*(10), 3183–3221. 4

Edmond, C., V. Midrigan, and D. Y. Xu (2018). How costly are markups? (document), 4, 4.3, 5, 5

Eggertsson, G. B., J. A. Robbins, and E. G. Wold (2018). Kaldor and Piketty's facts: The rise of monopoly power in the united states. Technical report, National Bureau of Economic Research. (document)

Epifani, P. and G. Gancia (2011). Trade, markup heterogeneity and misallocations. *Journal of International Economics 83*(1), 1–13. 4

Flynn, Z. (2019). Identifying productivity when it is a choice. *Working Paper*. 8, 4.4

Flynn, Z. and A. Gandhi (2018). Partial identification of production functions with flexible inputs. *Working Paper*. 1.2, 4.4

Foster, L., J. Haltiwanger, and C. Syverson (2008). Reallocation, firm turnover, and efficiency: Selection on productivity or profitability? *American Economic Review 98*(1), 394–425. (document), 7

Gandhi, A., S. Navarro, and D. Rivers (2017). On the identification of production functions: How heterogeneous is productivity? *Conditionally Accepted, Journal of Political Economy*. (document), 1.1, 2, 5.1, 7

Hall, R. E. (1988). The relation between price and marginal cost in US industry. *Journal of Political Economy 96*(5), 921–947. (document), 1.2

Hall, R. E. (2018). New evidence on the markup of prices over marginal costs and the role of mega-firms in the US economy. 1.2

Hoch, I. (1958). Simultaneous equation bias in the context of the Cobb-Douglas production function. *Econometrica*, 566–578. 5

Hsieh, C.-T. and P. Klenow (2009). Misallocation and manufacturing tfp in china and india. *Quarterly Journal of Economics 124*(4), 1403–1448. (document)

Karabarbounis, L. and B. Neiman (2013). The global decline of the labor share. *Quarterly Journal of Economics 129*(1), 61–103. (document)

Klette, T. J. and S. Kortum (2004). Innovating firms and aggregate innovation. *Journal of Political Economy 112*(5), 986–1018. 4.2

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

Marschak, J. and W. Andrews (1944). Random simulataneous equations and the theory of production. *Econometrica 12*, 143–205. 5

Melitz, M. J. and G. I. Ottaviano (2008). Market size, trade, and productivity. *The Review of Economic Studies 75*(1), 295–316. (document)

Olley, S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*, 1263–1297. 2

Peters, M. (2018). Heterogeneous mark-ups, growth and endogenous misallocation. (document), 4, 4.2, 4.2, E, E

Syverson, C. (2004). Market structure and productivity: A concrete example. *Journal of Political Economy 112*(6), 1181–1222. (document), 4, 7

Syverson, C. (2019). Macroeconomics and market power: Facts, potential explanations, and open questions. *Brookings Center on Regulation and Markets*. 1

Traina, J. (2018). Is aggregate market power increasing? Production trends using financial statements. *Stigler Center New Working Paper Series No. 17*. 4, 4.4, 5.1

A Proof of Theorem 1

Proof. Sufficiency:

Suppose that the only function Δ satisfying,

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]\right]+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta\left(v',k\right)dv'+\Delta\left(v,k\right)=0 \tag{78}$$

is $\Delta = 0$. Suppose, by way of contradiction, that there are two production functions in the identified set such that,

$$\int\int|f^1\left(v,k\right)-f^0\left(v,k\right)|dvdk\neq 0. \tag{79}$$

For a given flexible input elasticity, we can recover the production function using the proxy structure:

$$\begin{aligned} f\left(v,k\right)=\mathbb{E}\left[q_t-\int_{\overline{v}}^{v_t}f_v\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]-\\ \mathbb{E}\left[q_t-\int_{\overline{v}}^{v_t}f_v\left(v',k_t\right)dv'|k_t=\overline{k},v_{t-1},k_{t-1}\right]+\int_{\overline{v}}^{v}f_v\left(v',k\right)dv' \end{aligned} \tag{80}$$

Because both f^1 and f^0 satisfy the proxy structure, we can expressed them as above. A production function that satisfies the proxy structure has constant returns to scale if

$$\begin{aligned} 1=f_k\left(v,k\right)+f_v\left(v,k\right)=\frac{\partial}{\partial k}\mathbb{E}\left[q_t-\int_{\overline{v}}^{v_t}f_v\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]\\ +\int_{\overline{v}}^{v}f_{vk}\left(v',k\right)dv'+f_v\left(v,k\right) \end{aligned} \tag{81}$$

Both f^1 and f^0 have constant returns to scale, so we can difference the above equations to get the following restriction on the difference between the two flexible output elasticities:

$$0=-\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\left(f_v^1-f_v^0\right)\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\left[f_v^1-f_v^0\right]dv'+f_v^1-f_v^0 \tag{82}$$

Define $\Delta\left(v,k\right) = f_v^1 - f_v^0$. Then the identification assumption $\Delta = 0$ implies $f_v^1 = f_v^0$. Because both have constant returns to scale, this implies that either $f_k^1 = f_k^0$, or that the two production functions are the same which is a contradiction of the assumption that they are different. There cannot be more than one production function in the identified set given the identification condition is true.

Necessity:

Suppose there exists a $\Delta\left(v,k\right)\neq 0$ for all $\left(v,k\right)$ such that,

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta\left(v',k_t\right)dv'|k_t=k,v_{t-1},k_{t-1}\right]\right]+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta\left(v',k\right)dv'+\Delta\left(v,k\right)=0 \tag{83}$$

then, if there exists a production function f in the identified set, the production function has constant returns to scale, or:

$$f_k(v,k)+f_v(v,k)=\frac{\partial}{\partial k}\mathbb{E}\left[q_t-\int_{\overline{v}}^{v_t} f_v(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]+\int_{\overline{v}}^{v} f_{vk}(v',k)\,dv' \\ +f_v(v,k)=1. \quad (84)$$

We can construct another production function $\widetilde{f}$ that satisfies the identifying assumptions in the following way:

1. Let $\widetilde{f}_v=f_v+\Delta$.
2. Applying the same proxy transformation,

$$\widetilde{f}=\int_{\underline{k}}^{k}\frac{\partial}{\partial k}\mathbb{E}\left[q_t-\int_{\overline{v}}^{v_t} f_v(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]dk+\int_{\overline{v}}^{v} f_v(v',k)\,dv' \\ -\int_{\underline{k}}^{k}\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]dk+\int_{\overline{v}}^{v}\Delta(v,k) \quad (85)$$

$$\implies \widetilde{f}=f-\int_{\underline{k}}^{k}\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]dk+\int_{\overline{v}}^{v}\Delta(v,k) \quad (86)$$

3. Differentiating gives

$$\widetilde{f}_k=f_k-\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta(v',k)\,dv' \quad (87)$$

$$\widetilde{f}_v=f_v+\Delta(v,k) \quad (88)$$

$$\widetilde{f}_v+\widetilde{f}_k=1-\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right] \quad (89)$$

$$+\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta(v',k)\,dv'+\Delta(v,k)$$

4. Because this must be true for all (v,k,v_{t-1},k_{t-1}) and the left hand side is only a function of (v,k), integrating both sides with respect to (v_{t-1},k_{t-1}) makes no difference so:

$$\widetilde{f}_v+\widetilde{f}_k=1-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t}\Delta(v',k_t)\,dv'|k_t=k,v_{t-1},k_{t-1}\right]\right] \\ +\int_{\overline{v}}^{v}\frac{\partial}{\partial k}\Delta(v',k)\,dv'+\Delta(v,k) \quad (90) \\ =1$$

Implying $\widetilde{f}$ has constant returns to scale. Hence, there are multiple production functions in the identified set, a contradiction.

□

B Linear-in-Parameters Example

More generally, production functions that are linear-in-parameters (like the Cobb Douglas and translog production functions) will satisfy the identification condition under a reasonable rank assumption. Suppose,

$$f(v,k) = r(v,k)^\top \theta,$$

for some known vector of functions $r(v,k)$. Define $\Delta_\theta = \theta - \theta'$. Assumption 1 can then be written as,

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t} r_v(v',k)dv'|k_t\right]\right]^\top \Delta_\theta + \Delta_\theta^\top \int r_{vk}(v',k)dv' \tag{91}$$

$$+r_v(v,k)^\top \Delta_\theta = 0 \tag{92}$$

By the constant returns to scale assumption, we have:

$$\Delta_\theta^\top \int_{\overline{v}}^{v} r_{kv}(v',k)dv' = -\Delta_\theta^\top \int_{\overline{v}}^{v} r_{vv}(v',k)dv' = \Delta_\theta^\top \left[r_v(\overline{v},k) - r_v(v,k)\right] \tag{93}$$

So:

$$-\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[\int_{\overline{v}}^{v_t} r_v(v',k)dv'|k_t = k, v_{t-1}, k_{t-1}\right]\right]^\top \Delta_\theta + \Delta_\theta^\top r_v(\overline{v},k) = 0 \tag{94}$$

From the fundamental theorem of calculus and because $[r_v(\overline{v},k) + r_k(\overline{v},k)]^\top \Delta_\theta = 0$, we have the following condition,

$$-\mathbb{E}[\frac{\partial}{\partial k}\mathbb{E}\left[r(v_t,k)|k_t = k, v_{t-1}, k_{t-1}\right]]^\top \Delta_\theta = 0 \tag{95}$$

So, as long as the functions (of k) interacted with Δ_θ are not collinear the linear-in-parameters production function is identified. For example, if,

$$\mathbb{E}\left\{\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[r(v_t,k)|k_t = k, v_{t-1}, k_{t-1}\right]\right]\mathbb{E}\left[\frac{\partial}{\partial k}\mathbb{E}\left[r(v_t,k)|k_t = k, v_{t-1}, k_{t-1}\right]\right]^\top\right\},$$

is invertible, then the production function is identified.

C Illustrating with Linear-in-Parameters Production Functions

Suppose the production function can be written as,

$$f = r(v,k)^{\top}\theta. \tag{96}$$

For a known vector of functions $r(v,k)$. The model can be estimated in the following way:

1. Regress q_t on some flexible transformation of (v_t, k_t) to estimate $\phi(v_t, k_t) = \mathbb{E}[q_t|v_t, k_t]$.
2. For a given guess of θ,

$$\phi(v_t, k_t) - r(v_t, k_t)^{\top}\theta = a_t \tag{97}$$

3. Regress a_t on (transformations of) a_{t-1},

$$a_t = g(a_{t-1}) + \eta_t \tag{98}$$

4. Let $h(k)$ be a sufficient number of linearly-independent transformations of k. Solve the moment equation using GMM,

$$\frac{1}{n}\sum_{it} h(k_{it})\eta_{it}(\theta) = 0$$

$$\frac{1}{n}\sum_{it} r(v_{it}, k_{it})(RTS_{it}(\theta) - 1) = 0 \tag{99}$$

D Solution to the Monte Carlo Dynamic Programming Problem

To solve the dynamic programming problem, write out the first order conditions for (V, X, Q_E),

$$(V): \quad W_V = \lambda \times A\theta V^{\theta-1}K^{\gamma} \tag{100}$$

$$(X): \quad W_X = \beta\mathbb{E}[M_K(A\eta, \delta K + X)] \tag{101}$$

$$(Q_E): \quad (\eta+1)Q_E^{\eta}\mathbb{E}[\exp((\eta+1)\epsilon)] = \lambda \tag{102}$$

The solution to the static part of the optimization problem is found by collapsing the (Q_E, V) first order conditions,

$$W_V = (\eta+1)\left(AV^{\theta}K^{\gamma}\right)^{\eta}\mathbb{E}[\exp((\eta+1)\epsilon)] \times A\theta V^{\theta-1}K^{\gamma} \tag{103}$$

$$\implies V(A,K) = \left\{\frac{W_V \times A^{-\eta-1}K^{-\gamma(\eta+1)}}{(\eta+1)\theta\mathbb{E}[\exp((\eta+1)\epsilon)]}\right\}^{\frac{1}{\theta(\eta+1)-1}} \tag{104}$$

We can solve for the choice of X by using the envelope theorem and the first order conditions to write,

$$M_K(A,K) = \frac{\gamma(\eta+1)}{K} \times Q_E^{\eta+1}(A,K)^{\eta+1}\,\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right] + \beta\delta\mathbb{E}\left[M_K\left(A\eta, \delta\times K + X\right)\right] \tag{105}$$

$$= \frac{\gamma(\eta+1)}{K} \times Q_E^{\eta+1}(A,K)^{\eta+1}\,\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right] + \beta\delta W_X \tag{106}$$

Defining $K' = \delta \times K + X$ we can the X first order condition as,

$$(1-\beta\delta)W_X = \beta\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right]\mathbb{E}\left[\frac{\gamma(\eta+1)}{K'} \times Q_E\left(A\nu, K'\right)^{\eta+1}\right] \tag{107}$$

Because,

$$Q_E(A,K) = A \times \left\{\frac{W_V \times A^{-\eta-1}K^{-\gamma(\eta+1)}}{(\eta+1)\,\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right]}\right\}^{\frac{\theta}{\theta(\eta+1)-1}} \times K^{\gamma} = A^{-(\eta+1)\frac{\theta}{\theta(\eta+1)-1}+1} \times \iota_0 K^{\iota_1}, \tag{108}$$

where ι_0 and ι_1 are just functions of the problem's parameters.

Then, the first order condition for X is,

$$(1-\beta\delta)W_X = \beta\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right]\gamma(\eta+1) \times \iota_0 K'^{\iota_1(\eta+1)-1} \times A^{-\frac{(\eta+1)}{\theta(\eta+1)-1}}\,\mathbb{E}\left[\nu^{-\frac{(\eta+1)}{\theta(\eta+1)-1}}\right]. \tag{109}$$

Because ν is log normal (recall that $\mathbb{E}\left[\exp(tX)\right]$ is the moment generating function of X so we simply evaluate the moment generating function of $\log\nu$ at the relevant value),

$$\mathbb{E}\left[\exp\left(-\frac{(\eta+1)}{\theta(\eta+1)-1} \times \log\nu\right)\right] = \exp\left(-\frac{(\eta+1)}{\theta(\eta+1)-1}\mu_\nu + \frac{\sigma_\nu^2}{2} \times \left(\frac{(\eta+1)}{\theta(\eta+1)-1}\right)^2\right) \tag{110}$$

In any case, we can now solve the first order condition for X in terms of the parameters of the problem which yields the not-so-beautiful but easy-to-solve expression,

$$K'^{\iota_1(\eta+1)-1} = \frac{(1-\beta\delta)W_X}{\beta\mathbb{E}\left[\exp\left((\eta+1)\epsilon\right)\right]\gamma(\eta+1)\,\iota_0\exp\left(-\frac{(\eta+1)}{\theta(\eta+1)-1}\mu_\nu + \frac{\sigma_\nu^2}{2} \times \left(\frac{(\eta+1)}{\theta(\eta+1)-1}\right)^2\right)}A^{\frac{(\eta+1)}{\theta(\eta+1)-1}} \tag{111}$$

E Robustness Checks with Classification Methods

For robustness, we consider using classification methods from the machine learning literature to allow for more flexible production functions. We suppose that there are groups of firms within each industry with the same production function, but that these groups are unknown to us so we must learn them from the data unlike in the above results where we group the firms ex-ante.

The first method we consider is to group firms (within an industry) with similar input cost shares. Define the share of services in variable expenses as (SGA/OPEX). Economic theory suggests that firms with similar output elasticities will have similar fractions of their variable spending attributed to each of the variable inputs. We build a regression tree on the logit transformation of (SGA/OPEX) using firm dummies and year dummies as the covariates and leave-one-out least squares cross validation to score the model.

Regression trees work by splitting each set of covariates (firm dummies and year dummies) into binary groups and then estimating the leave-one-out least-squares cross validation metric using those group dummies as covariates. They iteratively split the sample in a hierarchical fashion to maximally reduce the score. They stop splitting when they fail to reduce the cross-validation score. Within each group created by this regression tree, we estimate a different production function using the method we develop in this paper.

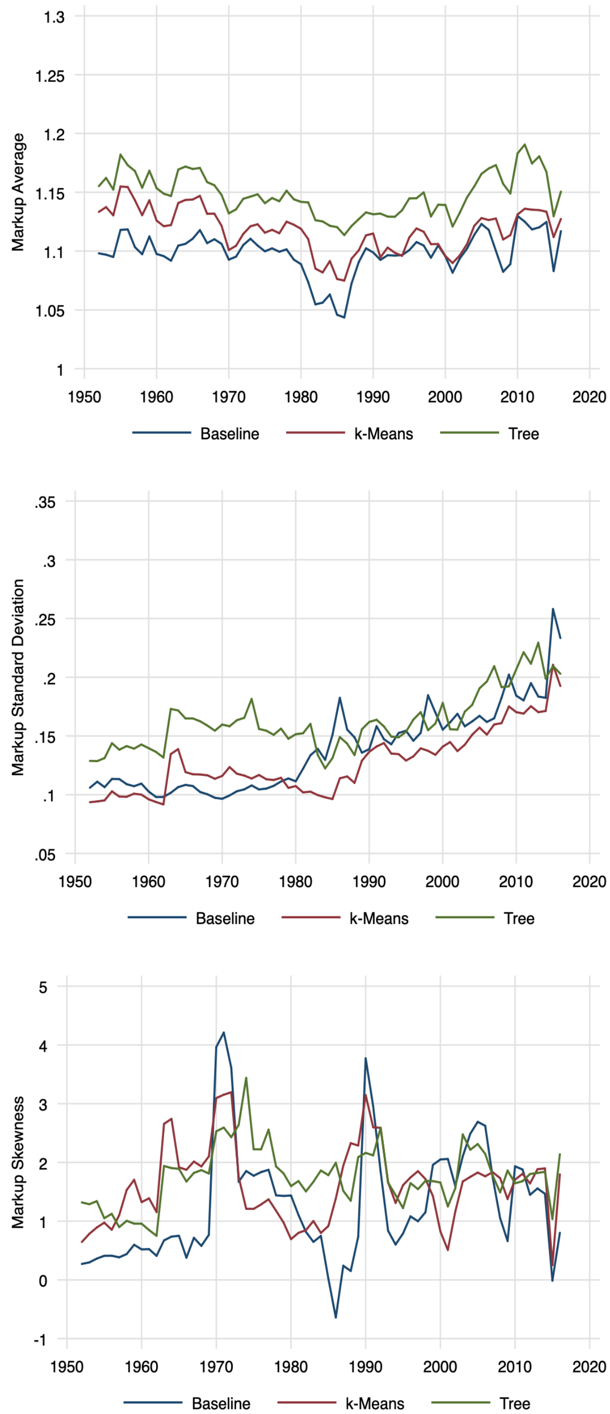
The second method we consider is k-means clustering, which allows us to group firms that are similar across a vector of characteristics. Within each industry, we group firms for the vector of (COGS/SALES, SGA/SALES, SALES). We use k=3 to form three groups within each industry and estimate the production function within each group.

Table 5 repeats the earlier exercise of evaluating estimators with the partial identification bounds. Although largely similar, the classifiers seem to decrease inadmissibly low markups, but increase inadmissibly high markups. The overall size of these violations remains small.

Figures 8 and 9 show the moments and model-implied losses for the CD-OPEX specification across classifiers. The takeaway is the same as Table 5 – these additional methods do not seem to pick up any notable heterogeneity not already covered by the flexibility of our proxy approach.

Figures 10 and 11 repeat this exercise but for the TL-COGS specification. The overall losses are considerably higher. Output losses are now about 4 percent (up from 1), and labor share losses are now about 21 percent (up from 11).

Figure 8: Level and Dispersion Stability, CD-OPEX



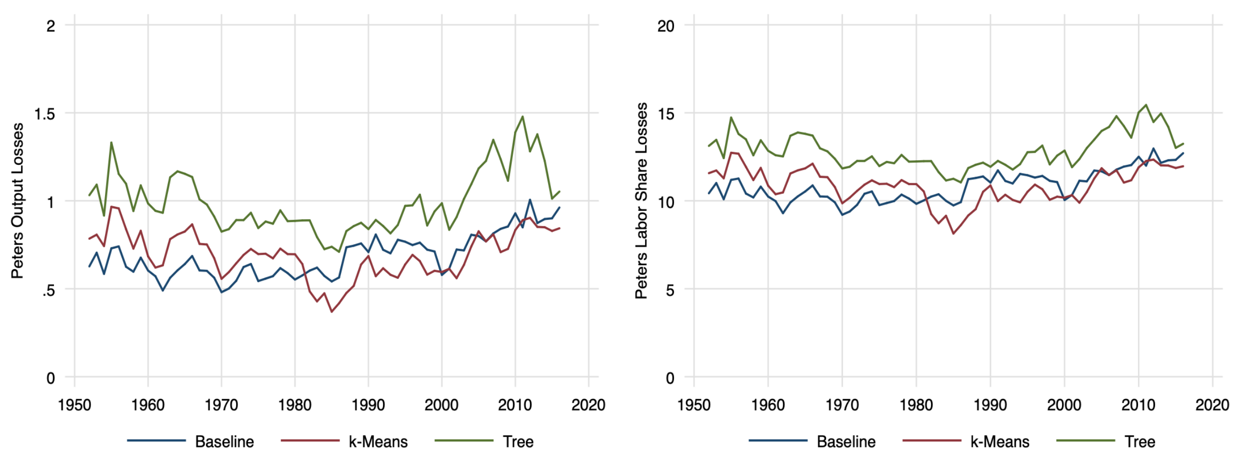
Source: Data from Compustat, estimates from authors' calculations.

Table 5: Specification Error using Bounds

Specification		Too Low, Raw Percent			Too High, Raw Percent		
Form	Flex	Base	k-Means	Tree	Base	k-Means	Tree
CD	OPEX	0.16	0.10	0.08	0.03	0.17	0.36
CD	COGS	0.31	0.37	0.22	0.41	0.48	0.60
TL	OPEX	0.18	0.10	0.16	0.62	0.65	0.63
TL	COGS	0.41	0.23	0.32	0.00	0.00	0.04

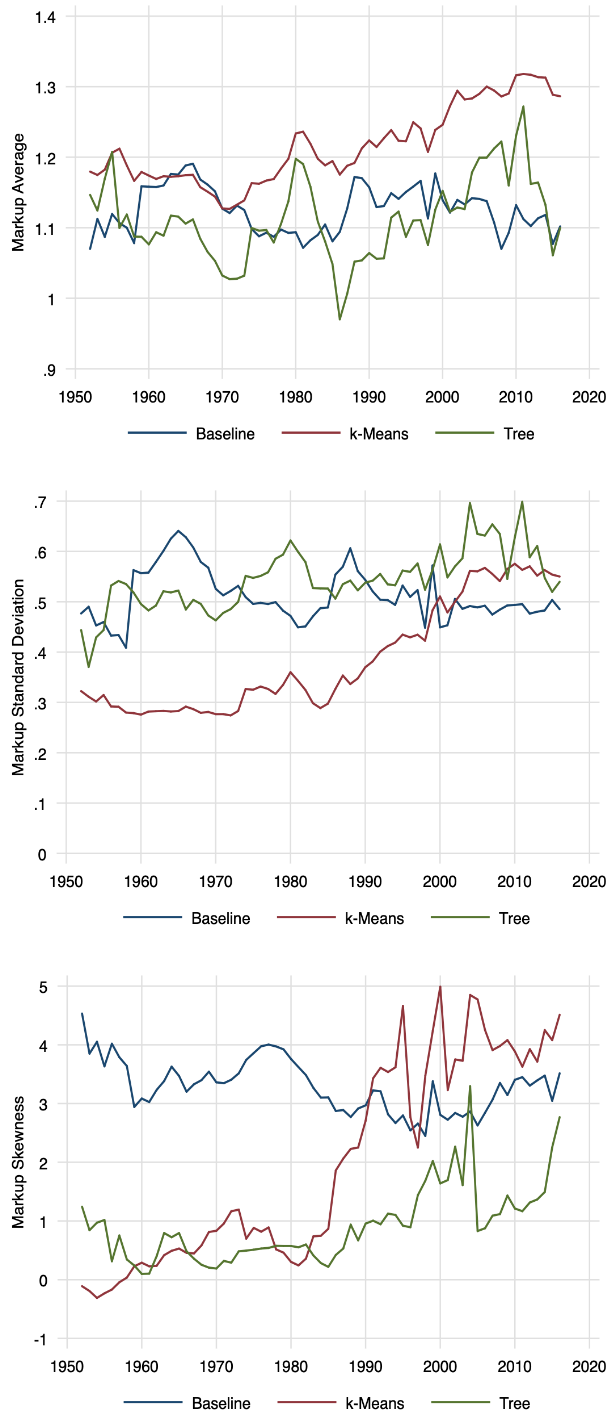
		Too Low, Average Violation			Too High, Average Violation		
Form	Flex	Base	k-Means	Tree	Base	k-Means	Tree
CD	OPEX	0.02	0.01	0.01	0.00	0.02	0.04
CD	COGS	0.04	0.04	0.06	0.11	0.13	0.12
TL	OPEX	0.02	0.01	0.07	0.06	0.08	0.09
TL	COGS	0.13	0.05	0.16	0.00	0.00	0.00

Figure 9: Macroeconomic Implications Stability, CD-OPEX



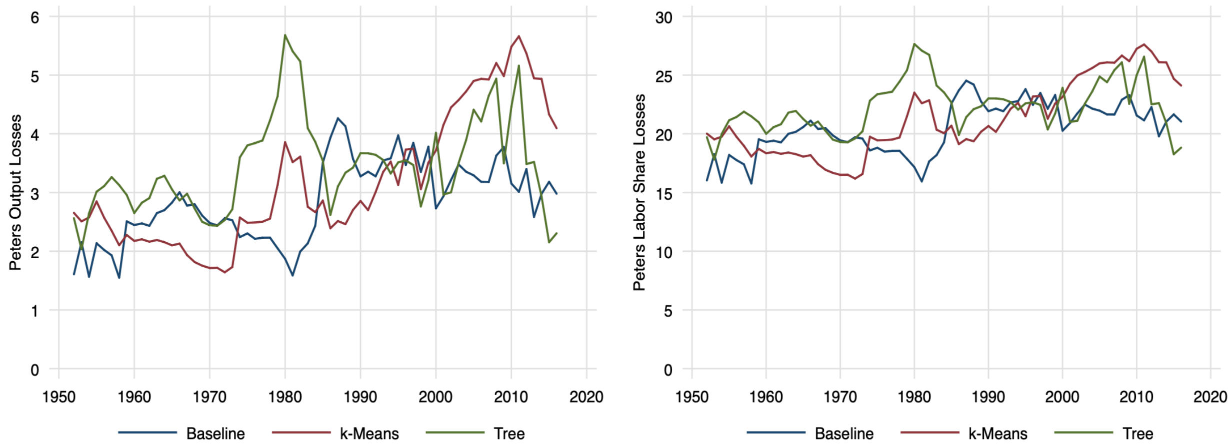
Source: Data from Compustat, estimates from authors' calculations using the sufficient statistics of Peters (2018).

Figure 10: Level and Dispersion Stability, TL-COGS



Source: Data from Compustat, estimates from authors' calculations.

Figure 11: Macroeconomic Implications Stability, CD-OPEX



Source: Data from Compustat, estimates from authors' calculations using the sufficient statistics of Peters (2018).

F Note on Revenue Data

Like in the vast majority of production settings, our data contain information on sales divided by price indices, not quantities directly. In this short section, we make explicit the (usually implicit) set of assumptions that imply identification of the *physical* production function with this measure of output.

Let log price be p. We can write the production equation as

$$p_t + q_t = f(v_t, k_t) + a_t + p_t + \epsilon_t \tag{112}$$

Define $\omega_t = a_t + p_t$, and modify the typical proxy assumptions about ω_t as follows:

Assumption 6. ω_t is a Markov process:

$$\omega_t = g(\omega_{t-1}) + \eta_t \tag{113}$$

Assumption 7. Future input choices have no information about future ω_{t+1} given current input choices. In particular,

$$g(\omega_t) = \mathbb{E}[g(\omega_t) | k_t, v_t] + u_t \tag{114}$$

$$\mathbb{E}[u_t | k_{t+1}, k_t, v_t] = 0. \tag{115}$$

In words, inputs span all systematic elements that determine ω_t, and future capital choice is unrelated to deviations from the projection of ω_t on inputs. Assumption 7 is a weaker form of the proxy assumption which assumes that $\omega_t = \omega(k_t, v_t)$. It allows for price to vary based on unobserved state variables so long as those state variables are either controlled for given (k_t, v_t), or have only idiosyncratic components that do not affect investment decisions made at time t.

Given Assumptions 6 and 7, we can write,

$$p_t + q_t = f\left(v_t, k_t\right) + \mathbb{E}\left[g\left(\omega_{t-1}\right)|k_{t-1}, v_{t-1}\right] + u_t + \eta_t + \epsilon_t. \tag{116}$$

We can then proceed applying the same estimation and identification argument developed in the main text. While we cannot identify physical productivity with this argument, we can identify the flexible input elasticity that we need to identify markups. The measure of productivity that does come out of the model may be viewed as a "quality-adjusted" productivity because it is the product of price and productivity, $a_t + p_t$.