Identifying productivity when it is a factor of production

Zach Flynn*

June 29, 2019

Abstract

Productivity is typically modeled as an exogenous plant characteristic, but the people who run and work manufacturing plants make choices that affect how productive they are. This article develops a method to partially identify statistics of the productivity distribution when productivity is a choice. The method is based on a monotone comparative static result developed in a general economic model and does not require instrumental variable or timing assumptions. I use the method to study the effect of restructuring on productivity in the electricity generation industry.

*Email: zlflynn@gmail.com. I thank Amit Gandhi, Alan Sorensen, Ken Hendricks, Enghin Atalay, Jack Porter, Xiaoxia Shi, Daniel Quint, Michael Dickstein, Nathan Yoder, Andrea Guglielmo, James Traina, Seth Benzell, participants at seminars at the University of Wisconsin - Madison, University of California Davis, Louisiana State University, the Federal Trade Commission, the Brattle Group, Chad Syverson (the editor), and two anonymous referees for comments and criticism that improved this article.

1 Introduction

While productivity is typically modeled in empirical work as an exogenous shock plants react to[1], plants make choices that affect their productivity. Plant owners choose the plant's technology. Workers choose whether to work hard or slack off. Managers choose whether to closely monitor workers or to let things slide. Plants are comprised of *people* who make choices daily in response to the incentives they face that affect the plant's productivity. This motivates the development of a general empirical strategy to deal with endogenous productivity.

I propose a method to partially identify productivity when it is a factor of production under the plant's control. I build a general economic model which encapsulates a broad class of common models in the literature. I then prove a comparative static result that implies a plant's productivity, output, and capacity choices are increasing in the latent, unobserved variables that determine its residual demand curve and cost function. Modeling these latent variables as a Markov process, I derive a restriction on the statistical relationship between productivity, output, and plant capacity using the comparative static result. Requiring the productivity distribution to satisfy this restriction allows us to meaningfully partially identify interesting economic parameters because some ex-ante reasonable production functions imply productivity distributions that do not satisfy the restriction. In particular, this identification strategy can be used to bound coefficients in regressions where productivity is the dependent variable. For standard parametrizations of the production function, these bounds can be computed using linear programming methods so they are simple to compute. I use this strategy to estimate bounds on the effect of restructuring in the electricity generation industry on power plant productivity.

The basic problem of identifying the production function is that input choice is correlated with productivity. This problem applies whether productivity is endogenously chosen or it is an exogenous parameter of the plant's production function. The identification problem has been discussed in a large and old literature from Marschak and Andrews (1944) and Griliches and Mairesse (1995) to the modern "proxy" approach to structural production function estimation developed in Olley and Pakes (1996), Levinsohn and Petrin (2003), Ackerberg, Caves, and Frazer (2015), and Gandhi, Navarro, and Rivers (2019) based on the models of firm and industry dynamics developed in Jovanovic (1982) and Hopenhayn (1992)[2].

[1]Most empirical papers that estimate productivity use models that suppose either productivity is exogenous or investment in productivity can be controlled for by observables, see the proxy approach of Olley and Pakes (1996) and Doraszelski and Jaumandreu (2013). But the idea that productivity depends on unobserved factors of production appears at least as early as Griliches and Jorgenson (1967). They argued that if the factors of production, both physical and intangible, were fully accounted for, there would be little left of the productivity residual — or, as they memorably called it, "the measure of our ignorance". The substantial identification problems this view of productivity introduces are the subject of this article.

[2]The proxy approach has been applied to a wide-range of empirical problems. These applications include studying the effects of trade liberalization (Pavcnik, 2002) and the effect of restructuring on the telecommunications equipment industry (Olley and Pakes, 1996) and, more recently, to estimate markups (De Loecker and Warzynski, 2012; De Loecker, Eeckhout, and Unger, 2019; Flynn, Gandhi, and Traina, 2019). See Syverson (2011) for a broader survey of applications.

The primary method that exists in the prior literature for estimating endogenous productivity models is developed in Doraszelski and Jaumandreu (2013) and De Loecker (2013). While the original proxy models were purely exogenous productivity models, they propose a modification of the proxy model that allows productivity to be endogenous when it is controlled by *observed* choices[3] that are assumed to be chosen before any unobserved shock to productivity. Relative to this work, I allow for unobserved investment in productivity and do not assume when this choice is made. I also do not make the assumption that the only unobserved state variable differentiating plants is productivity or that productivity is a function of observables, the namesake assumption of the proxy approach to production function estimation.

I develop the model and identification result in Section 2.

I show how to make inference on coefficients in regressions where log productivity is the dependent variable in Section 3. Regressions of productivity on policy variables and plant characteristics are a standard statistic of interest in the productivity literature, especially when the goal is to evaluate the effect of a policy on productivity or to learn which kinds of plants are more productive than others. The inference strategy is Bayesian and based on results from Kline and Tamer (2016). I formulate the bounds on the coefficients in the regression as the values of two linear programming problems. Inference on the identified set is made by drawing from the posterior of a set of reduced-form parameters and re-solving the linear programs repeatedly.

I discuss why I pursue the comparative static approach instead of an extension of the proxy or instrumental variable approaches in Section 4.

In Section 5, I demonstrate the bounds are narrow enough to be useful in practice by using them in practice. I bound the effect of *restructuring* in the electricity generation industry on power plant productivity. Historically, electricity prices were set by state-run public service commissions on the basis of the costs incurred in producing the electricity. In the mid 1990's to early 2000's, some US states restructured the industry by allowing electricity prices to be set by markets instead of by public service commissions. Other states maintained regulated pricing. If prices are set on the basis of costs, there is less of an incentive to reduce those costs so ending regulated pricing might have encouraged power plants to choose to be more productive, making this problem a good example of the empirical relevance of allowing for endogenous productivity.

But there are other effects of restructuring that encourage power plants to reduce their productivity. Electric utilities were originally integrated across the three stages of electricity production: generation, transmission, and retail to end consumers. Restructuring forced utilities to disintegrate so that transmission, the naturally monopolistic stage of electricity production, could be regulated while allowing markets to set prices in the generation stage. There may have been efficiencies from integration that were lost with restructuring. In addition, the Averch and Johnson (1962) effects of rate-of-return regulation (capital investment is prized over spending on other inputs) and increased competition among the utilities post-

[3] Van Biesebrock (2003) also studied observed technology choice and how it affected productivity where he explicitly observes automobile plants adopting different technologies.

restructuring may also contribute to lower productivity choice. I discuss these effects in more detail in Section 5.1.

I study whether restructuring incentivized power plants to increase their "fuel productivity". A power plant has a higher fuel productivity if it produces more output for a given amount of fuel, holding its nonfuel inputs constant. This multi-factor productivity measure appears in a natural model of electricity production I develop in Section 5.3. I find restructuring caused power plants to lower their fuel productivity by between 1.12% and 2.87%. Because productivity is a choice made at a cost, if a policy lowers productivity, it does not necessarily mean that it reduces welfare.

This article is not the first to analyze the effect of restructuring on productivity. The closest empirical paper is Fabrizio, Rose, and Wolfram (2007). They estimate the effect of restructuring on conditional input demand equations and use a proxy for demand (total electricity sales in a state, a measure of market size) to instrument for output. But demand is only uncorrelated with productivity if productivity is exogenous. Otherwise, demand affects productivity choice like it affects the choice of any other factor of production. So their model implicitly assumes productivity is exogenous. I establish results about the effect of restructuring on productivity that allow productivity to be endogenously adjusted by power plants in response to the new incentives offered by the policy. I discuss the difference between the measure of productivity used in Fabrizio, Rose, and Wolfram (2007) and the one used in this article as well as other differences between the two papers in Section 5.3.1.

Aside from the specific empirical result, the empirical application establishes that the bounds meaningfully restrict the parameters of interest in real applications.

2 Partial identification in a general model of production using a comparative static result

I develop an identification strategy that works when plants choose their productivity but does not depend delicately on a specific model of how they do so. The strategy is based on a monotone comparative static result derived from a general economic model. The result holds when productivity is a static decision[4] and when it is chosen with dynamic considerations[5], when plants are capacity-constrained and when they are unconstrained, when competition is perfect and when it is imperfect, and when productivity is a choice and when it is exogenous.

To give an outline of the identification strategy: I first show via monotone comparative static methods that output, capacity, and productivity choice are increasing in the unobserved state variables that determine a plant's cost function and its (residual) demand curve (Section 2.1). I pair this result with a statistical model of these unobserved state variables (Section 2.2). This pairing leads to the result that productivity is *positively associated* with output and

[4]For example, productivity is a static choice when it is determined primarily by effort as in principal-agent models of production or by quick-to-change logistical decisions about how production is organized like scheduling when certain people work or which clients a salesperson calls.

[5]For example, when productivity is a stock of capability like better machines or people.

capacity in the sense of Esary, Proschan, and Walkup (1967)[6]. I use this result to partially identify the production function and the productivity distribution (Section 2.3).

Throughout the article, I will use the following notation for the plant's production function. Let lowercase variables be in logs when uppercase variables are levels (for example, $\log Q = q$). Let Q be output, Z be a vector of L variable inputs, K be a capital or capacity input, A be total factor productivity, and F is the production function. Productivity is input neutral,

$$Q = F(Z, K)\, A \implies q = f(z, k) + a. \tag{1}$$

My ultimate goal is to identify coefficients in regressions of log productivity (a) on policy variables and plant characteristics (call these x),

$$a_t = x_t^\top \beta + u_t, \tag{2}$$

To do so, I first need to identify productivity so that I can perform the regression. Because $a_t = q_t - f(z_t, k_t)$ that requires identifying the log production function f.

I will show the production function is identified in the following set,

$$\begin{aligned} \mathcal{F} = \{ f \in \text{increasing functions}: \ &\text{for all functions } \phi_1 \text{ and } \phi_2 \\ &\text{that are increasing in all arguments}, \\ \operatorname{cov}\left[\phi_1 (q - f(z,k), q, k), \phi_2 (q - f(z,k), q, k)\right] &\geq 0 \} . \end{aligned} \tag{3}$$

Or, equivalently, a production function f is included in the identified set if it is increasing and the implied distribution of (a, q, k) is such that (a, q, k) is a positively associated random vector in the sense of Esary, Proschan, and Walkup (1967).

2.1 The model and a comparative static result

Plants choose their output Q, their capacity or capital K, and their productivity A to maximize discounted profits for a sequence of discount factors $\beta_t \geq 0$.

I describe each element of the model and then introduce the profit maximization problem the plant solves:

- $P(Q, \xi)$ gives the (residual) demand function for each plant. ξ varies across plants and indexes how the (residual) demand function varies across plants.
- $C\left(\frac{Q}{A}, K, W\right)$ is the variable cost function. Let Z be the variable inputs and W be input prices (which may vary by plant). When the plant is capacity-constrained, the cost function is,

$$C\left(\frac{Q}{A}, K, W\right) = \min W^\top Z \quad \text{ST:} \quad F(Z, K) \geq \frac{Q}{A} \text{ if } Q \leq K \tag{4}$$

$$= \infty \text{ if } Q > K. \tag{5}$$

[6]Positively associated random vectors have the property that the covariance of any two functions of the random vector that are increasing in all their arguments is positive.

I will prove two comparative static results: one for when K is capital and the plant is not capacity-constrained and another for when K is capacity.

- $M(A_t, A_{t-1}, \lambda_t)$ gives the cost of choosing a certain level of productivity, given the plant's past productivity. λ_t indexes how the cost of building productivity varies across plants.
- $G_t(K_{t+1}, K_t)$ gives the cost of building a certain level of capital or capacity given the current level in period t. The dependence on past capacity allows for it to be more expensive to go from low to high capacity than from medium to high capacity.
- For simplicity, I assume plants believe they have perfect foresight[7].

Plants solve the following problem:

$$\max_{K,A,Q} \sum_{t=1}^{T} \beta_t \times \left[\underbrace{P(Q_t, \xi_t)\, Q_t}_{\text{Revenue}} - \underbrace{C\left(\frac{Q_t}{A_t}, K_t, W_t\right)}_{\text{Production costs}} - \underbrace{M(A_t, A_{t-1}, \lambda_t)}_{\text{Technology cost}} - \underbrace{G_t(K_{t+1}, K_t)}_{\text{Capital adjustment costs}} \right]. \quad (6)$$

They make these choices subject to capacity (if capacity-constrained) and nonnegativity constraints,

$$0 \leq Q_t \leq K_t, \quad A_t \geq 0. \quad (7)$$

The model nests the basic proxy model developed in Olley and Pakes (1996). In the context of this model, the Olley and Pakes (1996) proxy model assumes productivity is *very* expensive to adjust (exogenous productivity)[8], ξ_t does not vary across plants (scalar unobservable), and there are no capacity constraints (capacity constraints would violate the proxy assumption of monotonicity of input demand in productivity for fixed capacity[9]).

2.1.1 Giving meaning to the parameters (ξ, λ)

I now put structure on the parameters of the plant's decision problem. My first task is to give the parameters ξ and λ a scalar meaning as opposed to indexing an entirely general function space. Only then, can we say something about what the demand curves of high ξ plants look like compared to low ξ plants. What this does is allow us to develop comparative

[7] Relaxing this assumption for risk-neutral plants with uncertainty about their future demand or cost of productivity is as straightforward as re-interpreting (ξ, λ) as differences in beliefs about those demand and costs.

[8] Exogenous productivity models are nested within the model by making it very expensive for the plant to adjust its productivity. For example, set $M(A_t, A_{t-1}, \lambda_t) = [A_t - \lambda_t]^{2\times\text{exponent}}$ where "exponent" is a very large integer (this M satisfies Assumption 1). Then, λ_t is the exogenous productivity sequence. Exogenous productivity models are endogenous productivity models where it is infinitely costly to adjust productivity.

[9] To see this, consider the Levinsohn and Petrin (2003) proxy: flexible inputs. The proxy model argues that $z_t = z_t(a_t, k_t)$ is increasing in a_t. For fixed capacity, increasing productivity eventually causes capacity constraints to bind. When capacity constraints bind, further increasing productivity will cause the plant to cut back on its flexible input use to keep its output below its capacity.

statics about how changes in ξ and λ affect plant choices. Assumption 1 specifies precisely the meaning of the two parameters. In short, greater ξ means greater marginal revenue for a given output level and greater λ means a lower marginal cost of productivity.

Assumption 1. Marginal revenue is increasing in ξ,

$$\frac{\partial}{\partial \xi}\left[\frac{\partial P}{\partial Q}(Q,\xi)\,Q + P(Q,\xi)\right] \geq 0. \tag{8}$$

The marginal cost of productivity is decreasing in λ_t,

$$\frac{\partial M}{\partial A_t \partial \lambda_t}(A_t, A_{t-1}, \lambda_t) \leq 0, \tag{9}$$

and, the reduction in marginal cost from higher prior productivity is also greater the greater λ_t is,

$$\frac{\partial M}{\partial A_{t-1} \partial \lambda_t}(A_t, A_{t-1}, \lambda_t) \leq 0. \tag{10}$$

2.1.2 Marginal cost of the variable inputs does not decrease too fast

I next impose a standard structure on the cost function for the Z inputs.

First, I assume that the cost function for the Z inputs is (weakly) convex in the required output. This does not rule out increasing returns to scale. For example, in a Cobb Douglas production function where Z is a single input, $Q = Z^{\theta_z} K^{\theta_k} A$. The cost function, conditional on K and A choices, is convex if $\theta_z \leq 1$ even if $\theta_z + \theta_k > 1$[10]. What it rules out is that $\theta_z > 1$, or, more generally, that there is increasing returns to scale in Z for fixed (K, A).

Second, I assume that inputs are *normal*. Inputs are normal if their conditional input demand functions are increasing in output. The best argument for the normality of inputs is the behavior of non-normal inputs. Demand for non-normal inputs increases when the price of the input *increases*. It seems unlikely that the response to an increase in fuel prices will cause power plants to purchase more fuel, but that is what we have to believe if fuel were not a normal input. This assumption is based on the principle that increases in the price of things will, all else equal, decrease demand for them.

Lastly, I assume that to the extent K is capital not capacity, it increases the marginal product of the variable inputs. The reasonableness of this assumption will vary from application-to-application, but capital is likely complementary with variable inputs in many production problems. Better machines make people and materials more productive inputs. The Cobb

[10] Because productivity is also a factor of production, the returns to scale in this model also depend on the cost of productivity. It is not simply the sum of the output elasticities of the production function F.

Douglas production function satisfies this assumption.

Assumption 2 gives the structure I put on the plant's variable cost function.

Assumption 2. The cost function of the Z inputs is convex,

$$\frac{\partial^2 C}{\partial Q^2} \geq 0. \tag{11}$$

The Z-inputs are *normal*: their conditional factor demands are increasing in output.

$$\frac{\partial Z\left(\frac{Q}{A}, K, W\right)}{\partial Q} \geq 0. \tag{12}$$

To the extent K is capital not capacity, it increases the marginal product of the Z inputs,

$$\frac{\partial F\left(Z, K\right)}{\partial K \partial Z_\ell} \geq 0 \quad \forall \ell = 1, \ldots, L. \tag{13}$$

Given Assumption 2, I can derive a result I use to establish the main comparative static result. Lemma 1 establishes that Assumption 2 implies $\frac{\partial C}{\partial Q \partial K} C\left(\frac{Q}{A}, K, W\right) \leq 0$.

Lemma 1. Given Assumption 2, marginal cost is decreasing in capital,

$$\frac{\partial C}{\partial Q \partial K} C\left(\frac{Q}{A}, K, W\right) \leq 0. \tag{14}$$

Proof. See Appendix B. □

2.1.3 Convex investment costs for capital and productivity

G_t gives the cost of accumulating future capital (or capacity) as a function of past capital. A standard specification for G_t is,

$$G_t\left(K_{t+1}, K_t\right) = \text{constant}_t \times \left(K_{t+1} - K_t\right)^{\text{exponent}_t}. \tag{15}$$

I assume that capital faces (weakly) convex adjustment costs. To make larger capital investments in a fixed amount of time requires greater marginal expenditure because it is progressively more expensive to purchase larger or better machines and more expensive to finance large expenditures than small ones. For the above example of G_t, exponents greater than 1 correspond to convex adjustment costs. Convex adjustment costs are widely assumed in economic modeling because they capture that it is more expensive to invest a lot at once than to invest a little this year and a little next year (ignoring discounting).

Similarly, $M(A_t, A_{t-1}, \lambda_t)$ gives the cost of investing in productivity. I assume convex adjustment costs for productivity for the same reason I assume them for capital investment. Greater productivity requires progressively better machines, better workers and managers, or the agent to expend greater and greater effort to increase productivity by the same amount.

Returning to the nonparametric $G_t(K_{t+1}, K_t)$, I make the following mathematical assumption to capture the economic assumption of convex investment costs.

Assumption 3. The cross partial of $G_t(K_{t+1}, K_t)$ is negative,

$$\frac{\partial G_t}{\partial K_t \partial K_{t+1}}(K_{t+1}, K_t) \leq 0 \tag{16}$$

The cross partial of $M(A_t, A_{t-1}, \lambda_t)$ with respect to (A_t, A_{t-1}) is negative,

$$\frac{\partial M}{\partial A_t \partial A_{t-1}} \leq 0. \tag{17}$$

We can see that this assumption holds for $G_t(K_{t+1}, K_t) = \text{constant}_t \times (K_{t+1} - K_t)^{\text{exponent}_t}$ when investment costs are convex,

$$\frac{\partial G_t}{\partial K_t \partial K_{t+1}}(K_{t+1}, K_t) = -\text{constant}_t \times (\text{exponent}_t - 1) \times \text{exponent}_t \tag{18}$$

$$\times (K_{t+1} - K_t)^{\text{exponent}_t - 2} \leq 0. \tag{19}$$

More generally, Assumption 3 holds for functions $G_t = \widetilde{G}_t(K_{t+1} - K_t)$ and $M = \widetilde{M}(A_t - A_{t-1}, \lambda_t)$ where $\widetilde{G}_t$ is a convex function and $\widetilde{M}(\cdot, \lambda_t)$ is a convex function for any given λ_t. This establishes that Assumption 3 captures the economic assumption of convex adjustment costs.

2.1.4 The revenue maximization problem has no interior solutions

I have put structure on the cost side of the plant's problem (C, G, and M). I now add structure to the demand side. I assume that the revenue maximization problem has no interior solution. There is no "bliss point" for output ignoring the costs of producing it. Costs limit plant output not market power. While this may not be the case for some products, like pharmaceuticals, it seems likely to be the case for many other products which either have sufficient competition so that the residual demand curve is not too inelastic or that demand for the product itself is elastic enough. We can operationalize this economic assumption with the following restriction on the plant's (residual) demand curve:

Assumption 4. The product of marginal revenue and output is increasing in Q,

$$\frac{\partial}{\partial Q}\left[\text{MR}(Q, \xi)\, Q\right] \geq 0 \quad \text{for all } \xi. \tag{20}$$

This assumption is true in a wide variety of demand models. It is true when the plant's demand curve has a constant elasticity, when demand is logistic (as in standard discrete choice models), when plants take prices as given, and for any elastic demand curve where the (absolute value of the) elasticity of demand is increasing in price[11].

Theorem 1 links Assumption 4 to the economic restriction that the revenue maximization does not have an interior solution.

Theorem 1. Given Assumption 4, if there exists no Q such that $\mathrm{MR}(Q) = \mathrm{MR}'(Q) = 0$, then there is no interior solution to the revenue maximization problem.

Proof. See Appendix B. □

2.1.5 Main comparative static result

Given this structure, I derive a comparative static result in Theorem 2 relating the choice of productivity, capital or capacity, and output to the plant's underlying state variables. The form of the result is slightly different depending on whether K is capital or whether it is capacity, but in either case, I will show it can be used to partially identify the production function and statistics of the productivity distribution.

The result makes use of the Topkis (1978) theorem. The theorem gives conditions under which objective functions maximized over a lattice[12] have maximizers that are increasing in the parameters of the optimization problem. The conditions required are:

(1) The objective function is supermodular in the plants choices (each pairwise cross-partial derivative is positive) and

(2) The objective function has increasing differences in its choices and which ever parameter we would like to show the choices are increasing in (each cross-partial of one of the plant's choices and the parameter is positive).

[11] To see this,

$$\mathrm{MR}(Q,\xi)\,Q = \underbrace{P(Q,\xi)\,Q}_{\text{Revenue}} \times \underbrace{\left[\frac{P'(Q,\xi)}{P(Q,\xi)} \times Q + 1\right]}_{\text{Inverse demand elasticity plus one}}$$

$$\begin{aligned}\implies \frac{\partial}{\partial Q}\left[\mathrm{MR}(Q,\xi)\,Q\right] = {} & \mathrm{MR}(Q,\xi) \times \frac{\partial}{\partial Q}\left[\frac{P'(Q,\xi)}{P(Q,\xi)} \times Q\right] \\ & + \text{Revenue}(Q,\xi) \times \left[\frac{P'(Q,\xi)}{P(Q,\xi)} \times Q + 1\right]\end{aligned}$$

If demand is elastic, the second term is positive because the inverse demand elasticity is greater than -1. If the elasticity of demand is decreasing in price (increasing in absolute value), the first term is positive because the inverse demand elasticity is then increasing.

[12] A lattice (for real vectors) is a set where the element-wise minimization and maximization of any two elements in the set also belong to the set.

These conditions as well as the requirement that we are maximizing over a lattice are *cardinal*. They may hold for certain transformations of the choices and not others. In the proof of the theorem, I make use of such a transformation.

Theorem 2. Define $s = \left((\lambda_t)_{t=1}^{T}, (\xi_t)_{t=1}^{T}, K_0, A_0 \right)$ as the vector of state variables. Let $Y^T = (Y_1, \dots, Y_T)$ for any variable Y. Given Assumptions 1, 2, 3, and 4, then

(1) If K is capital, so that $Q = F(Z, K)\, A$,

(2) OR if K is capacity, so that $Q = \min\{F(Z)\, A, K\}$,

then $A_t\left(s, -W^T\right)$, $Q_t\left(s, -W^T\right)$, and $K_t\left(s, -W^T\right)$ are increasing functions of s for fixed W^T, and, if K is capacity, then $A_t(\cdot)$, $Q_t(\cdot)$, and $K_t(\cdot)$ are increasing in all arguments, including W^T

Proof. See Appendix B. □

Theorem 2 establishes that plants choose output, productivity, and capacity in the same way. When the plant is incentivized to produce more output, it is also incentivized to be more productivity and to choose greater capacity.

2.2 Markovian state variables and positive association

For this result to say something about the data, we need a statistical structure for the unobserved state variables $\widetilde{s} = \left(\xi^T, \lambda^T, -W^T, K_0, A_0\right)$[13]. I show a standard Markovian structure on the state variables implies a useful statistical result. I focus on the assumptions I use for the case when K is capacity because that is what I use in my application, but, at the end of the section, I briefly state how the argument should be modified for the case when K is capital.

I assume that the state vector $\widetilde{s}$ is a Markov process, a standard assumption in dynamic economic models. For example, in the proxy model of production, the state variable is productivity and it is assumed to be a first order Markov process. I make a similar assumption for the state variables in this model with Assumption 5.

Assumption 5. Define $\widetilde{s}_t = (K_0, A_0, \xi_t, \lambda_t, -W_t)$. State variables evolve according to the rule,

$$\widetilde{s}_{j,t} = \sigma_{j,t}\left(\widetilde{s}_{t-1}, \upsilon_{j,t}\right), \tag{21}$$

where $\sigma_{j,t}$ is an increasing function for all (j,t) where j indexes the elements of $\widetilde{s}_t$.

Define $\widetilde{s}^{\tau} = \cup_{t=0}^{\tau} \widetilde{s}_t$. $\widetilde{s}^{t-1}$ and $\upsilon_t = (\upsilon_{1,t}, \dots, \upsilon_{J,t})$ are independent.

[13] If there is only one unobserved state variable that varies across plants, as in the proxy model, then none of this structure is necessary. The identification result will hold regradless.

The assumption that $\sigma_{j,t}$ is increasing nests the standard Markov structure, $\widetilde{s}_{j,t} = \sigma_{j,t}\left(\widetilde{s}_{j,t-1}, \upsilon_{j,t}\right)$, where greater $\widetilde{s}_{j,t-1}$ implies a larger distribution of $\widetilde{s}_{j,t}$ (persistence) but adds a bit of generality by allowing the other lagged state variables to positively affect another state variable.

I next add structure to the shocks $(\upsilon_{1,t}, \ldots, \upsilon_{J,t})$. I assume they are *positively associated*.

Esary, Proschan, and Walkup (1967) define positive association in the following way:

Definition 1 (Positive association)**.** Let X be a random vector. If for any two increasing functions, ϕ_1 and ϕ_2, such that the required expectations exist,

$$\operatorname{cov}\left[\phi_1\left(X\right), \phi_2\left(X\right)\right] \geq 0, \tag{22}$$

then X is a positively associated random vector.

Positively associated random vectors have the following properties:

Property 1. Suppose that X and Y are both positively associated vectors and statistically independent of each other. Then, (X, Y) is also positively associated. As an implication, if all elements of a random vector X are independent, then X is positively associated because any scalar random variable is positively associated.

Property 2. If X is positively associated, a vector of increasing functions of X, $B\left(X\right) = \left[b_1\left(X\right), \ldots, b_N\left(X\right)\right]$, is also positively associated.

Positive association is a reasonable structure to put on the shocks. If all the shocks $(\upsilon_{1,t}, \ldots, \upsilon_{J,t})$ are independent of one another, a standard modeling assumption, then they are positively associated. More generally, positive association allows for general *ability* shocks introducing a positive dependence between the shocks[14]. If there really is only one unobservabed state variable that varies across plants as assumed in the standard Olley and Pakes (1996) and Levinsohn and Petrin (2003) proxy model, then positive association holds by definition because the covariance of increasing functions of a single random variable is always positive, so the structure is more general than the standard statistical assumption in the literature.

To see that the model allows for *ability* shocks consider the following example model for υ,

$$\upsilon_{j,t} = \epsilon_t + \eta_{j,t} \tag{23}$$

where $\eta_{j,t}$ is independent of $\eta_{k,t}$. All of the state variables are "good" for the plant in the sense that they make the plant more profitable. So what this structure encapsulates is the idea that there is some general "talent" or quality of the plant shock in addition to shocks that affect the specific variables. The covariance of the shocks is then, $\operatorname{cov}\left(\upsilon_{j,t}, \upsilon_{k,t}\right) = \operatorname{var}\left(\epsilon_t\right) \geq 0$. This motivates allowing $(\upsilon_{1,t}, \ldots, \upsilon_{J,t})$ to be positively associated and not just making the stronger assumption that they are independent.

[14]I call it an "ability" shock because all the state variables increase plant profits.

Assumption 6. $\upsilon_t = (\upsilon_{1,t}, \ldots, \upsilon_{J,t})$ is a positively associated vector.

Assumptions 5 and 6 allow me to develop a simple result applying the properties of positive association above. The result says that positive association is an absorbing state: once the state variables are positively associated, they will always be.

Lemma 2. Given Assumption 5 and 6, if $\widetilde{s}_{t-1}$ is positively associated, then $\widetilde{s}_t$ is positively associated.

Proof. See Appendix B. □

To close the statistical model, I assume that we have reached this absorbing state in the data. The only way to not be in this absorbing state would be if initial draws of the state variable were negatively related. If the initial state variable draws are, say, independent of each other, then we would always be in the absorbing state. I assume that at some point prior to the data the state variables were positively associated.

Assumption 7. $(\xi_0, \lambda_0, A_0, K_0)$ is positively associated.

I can then prove Theorem 3 which concludes that $\widetilde{s}$ is positively associated.

Theorem 3. Given Assumption 5, 6, and 7, the state vector $\widetilde{s}$ is positively associated.

Proof. See Appendix B. □

For the case where K is capital, we need the additional assumption that $-W^T$ is independent of s because we need that s is positively associated conditional on W^T for the identification result to hold. Within the above structure, this amounts to the following restrictions:

$$\begin{aligned} s_t &= \sigma_{s,t}\left(s_{t-1}, \upsilon_{s,t}\right) && (24)\\ -W_t &= \sigma_{W,t}\left(-W_{t-1}, \upsilon_{W,t}\right) && (25)\\ \upsilon_{W,t} &\text{ is independent of } \upsilon_{s,t} && (26)\\ s_0 &\text{ is independent of } -W_0. && (27) \end{aligned}$$

These assumptions imply that s conditional on W^T is positively associated.

2.3 Identified set

Positive association is a useful structure for the problem because of Property 2. It is an ordinal property[15]: increasing functions of random vectors that are positively associated are themselves positively associated. Because we have established that the plant's choices $A_t\left(s,-W^T\right)$, $K_t\left(s,-W^{-T}\right)$, and $Q_t\left(s,-W^T\right)$ are increasing in $\widetilde{s}=\left(s,-W^T\right)$ and $\widetilde{s}$ is positively associated, we have that (A_t, K_t, Q_t) are positively associated. Because some distributions of A_t will satisfy this assumption and others will not, this result partially identifies the productivity distribution. I state the identification result formally in Theorem 4.

Positive association assumptions on the unobserved state variables are a general way of forming identification conditions from comparative static results. This method could be applied to other problems.

Theorem 4. Given the conclusions of Theorem 2 and 3 and data on (q, z, k), the production function f must satisfy the following conditions,

$$\begin{aligned}\mathcal{F}=\{f\in \text{increasing functions}:\ &\text{for all functions } \phi_1 \text{ and } \phi_2\\ &\text{that are increasing in all arguments },\\ &\operatorname{cov}\left[\phi_1\left(q-f\left(z,k\right),q,k\right),\phi_2\left(q-f\left(z,k\right),q,k\right)\right]\geq 0\}\,.\end{aligned}\tag{28}$$

If K is capital, then the covariance restriction must hold conditional on W^T.

Proof. By Theorem 3, $\widetilde{s}$ is positively associated. By Theorem 2, $a\left(\widetilde{s}\right)$, $q\left(\widetilde{s}\right)$, and $k\left(\widetilde{s}\right)$ are all increasing functions. By Property 2, increasing functions of positively associated random variables are positively associated. Therefore, (a, q, k) are positively associated. □

I will show that restricting the production function to this identified set implies surprisingly narrow bounds on common statistics of interest in the productivity literature.

3 Inference on statistics of the productivity distribution

In this section, I develop a practical method to use the partial identification result from Section 2 to form bounds on a common class of parameters of interest.

[15] Unlike other forms of positive dependence between random variables. For example, Manski and Pepper (2000) propose monotone instrumental variable assumptions which suppose the following form of positive dependence between covariates and the residual: $\partial\mathbb{E}\left[x_1|x_2\right]/\partial x_2\geq 0$. But this assumption does not necessarily imply $\partial\mathbb{E}\left[\phi\left(x_1\right)|x_2\right]/\partial x_2\geq 0$ if ϕ is increasing. Suppose $x_1\sim N\left(x_2, 2x_2^{-1}\right)$ conditional on x_2. Clearly, $\mathbb{E}\left[x_1|x_2\right]=x_2$ is increasing in x_2, but $\mathbb{E}\left[\exp\left(x_1\right)|x_2\right]=\exp\left(x_2+\frac{1}{x_2}\right)$ is decreasing for small x_2. Positive association is not stronger or weaker than monotone instrumental variable assumptions.

Throughout, I focus on making inference on coefficients in a regression of productivity on plant-level characteristics and/or policy variables instead of directly identifying the production function itself. To make the problem concrete, I consider estimating bounds on the parameter τ in the linear (instrumental variable) regression where log productivity is the dependent variable,

$$a_t = x_t^\top \beta + u_t \tag{29}$$
$$\mathbb{E}\left[v_t u_t\right] = 0 \tag{30}$$
$$\mathbb{E}\left[v_t x_t^\top\right] \text{ is left-invertible} \tag{31}$$
$$\tau = \rho^\top \beta, \tag{32}$$

where ρ is a known vector, x is a vector of observed covariates, and v is a vector of instruments for x ($v = x$ in regression problems)[16]. For example, τ could be a difference-in-difference or instrumental variable estimate of the effect of a policy on productivity or the average difference in productivity between two types of plants like the well-studied difference in productivity between exporters and non-exporters. All the parameters I estimate in my application have the same form as τ.

The advantage of focusing directly on forming bounds on the final parameter of interest can be seen by considering the alternative. The alternative is to form a confidence set for the production function, and then search over the production functions in that set, computing τ for each, a computationally intensive task. By focusing directly on the final parameter of interest, we can simplify computation of the identified set for τ. Focusing on inference on functions of partially identified parameter is a common strategy in the econometrics literature for this reason, see Kaido, Molinari, and Stoye (2016) and Bugni, Canay, and Shi (2016).

Let $\tau(y)$ be the estimate of τ when the dependent variable is y. Bounds on τ are, by definition,

$$\left[\min_{f \in \mathcal{F}} \tau\left(q - f\right), \max_{f \in \mathcal{F}} \tau\left(q - f\right)\right]. \tag{33}$$

So the bounds are the value of two constrained optimization problems.

3.1 Linear positive association: a convex identified set

It is difficult computationally to directly impose the restriction that (a, k, q) are positively associated to form these bounds. The identified set $\mathcal{F}$ is not convex. and nonconvex programming problems are difficult to solve. But there are well-established algorithms for convex programming like the log-barrier method that are guaranteed to converge in polynomial

[16] Presumably, these instruments do not include z, input use, or else we might be able to point identify the statistic of interest. For example, if $v = (z, x)$, then the statistic of interest is point identified. These are instruments for x which might be a policy variable and, say, plant fixed effects, not instruments that could be used to identify the production function.

time. So I first derive a relaxation of the full positive association assumption that is convex, making the bounds easier to compute.

I add one innocuous assumption to define the convex identified set. I assume there is no plant that cannot produce any output no matter how many inputs it uses. There is no plant with $A = 0$.

Assumption 8. There is a minimal level of productivity ($A \geq \underline{A} > 0$). Intuitively, there exists no plant that cannot produce any output no matter how many inputs it uses.

Given Assumption 8, I normalize the constant of the production function so that $\underline{A} = 1$.

I consider a convex relaxation of positive association which I call "linear positive association". It identifies the log production function in the following set,

$$\begin{aligned} \mathcal{F}_{\text{LPA}} = \{f : & Df \geq 0, q - f = a, a \geq 0, \\ & \operatorname{cov}\left[a\phi_1\left(q,k\right), \phi_2\left(q,k\right)\right] \geq 0 \\ & \text{ where } \phi_1 \geq 0 \text{ and } \phi_2 \text{ are increasing.}\} \,. \end{aligned} \tag{34}$$

Linear positive association is weaker than positive association (given Assumption 8) because a function f is in the identified set if it implies positive covariance between a certain class of increasing functions of (a, k, q) instead of all increasing functions of (a, k, q) as required in positive association. Importantly, the set identified by linear positive association has a nice property that does not hold for the set identified by positive association: $\mathcal{F}_{\text{LPA}}$ is convex.

Theorem 5. Suppose f^0 and f^1 are two elements of $\mathcal{F}_{\text{LPA}}$ then $f^\alpha = (1-\alpha)\, f^0 + \alpha f^1 \in \mathcal{F}_{\text{LPA}}$ for $\alpha \in [0, 1]$.

Proof. See Appendix B. □

Bounds on τ given that $f \in \mathcal{F}_{\text{LPA}}$ can then be found by solving a convex programming problem, which are much simpler problems to compute,

$$\text{max or min }_{f\in\mathcal{F}_{\text{LPA}}} \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t f\left(z_t\right)\right]. \tag{35}$$

3.2 Computing the bounds via linear programming for linear-in-parameters production functions

In practice, we often assume a parametric functional form for the production function that is linear in parameters, like the Cobb Douglas or translog production function. One of the main attractions of the linear positive association assumption in this case is that computing the bounds is as straightforward as solving two linear programming problems.

Suppose the production function can be written as,

$$f(z) = r(z)^\top \theta, \tag{36}$$

where $r(z)$ is a known vector of functions and θ is an unknown parameter vector. From the definition of $\mathcal{F}_{\text{LPA}}$, linear positive association identifies θ in a set determined by the intersection of infinitely many linear inequalities,

$$\begin{aligned}\Theta_{\text{LPA}} = \{\theta : & Dr\theta \geq 0, q - r(z)^\top \theta = a, a \geq 0, \\ & \operatorname{cov}[a\phi_1(q,k), \phi_2(q,k)] \geq 0 \quad \forall (\phi_1, \phi_2) \text{ where } \phi_1 \geq 0, D\phi_1 \geq 0,\ D\phi_2 \geq 0\}.\end{aligned} \tag{37}$$

While we could use all of these constraints in estimation, doing so would mean some of the constraints would have substantial identifying power via the assumed functional form of f. I propose using only a subset[17] of the constraints based on the functional form specified. This principle is commonly followed in empirical economics when estimating regressions. In regression problems of the form $y = \mathbb{E}[y|x] + \epsilon$, even though $\mathbb{E}[\epsilon|x] = 0$, when we use a linear regression to approximate $\mathbb{E}[y|x]$, we only use $\mathbb{E}[x\epsilon] = 0$ as the moment conditions in estimation. We do not form moment conditions by interacting ϵ with every possible transformation of x.

I follow this principle and only use increasing functions ϕ_1 and ϕ_2 that are similar in complexity to the parameter space we use for the production function itself. For example and because this is the functional form I use in my application, suppose the production function were Cobb Douglas,

$$r(z) = (1, z). \tag{38}$$

Following the principal that the functions (ϕ_1, ϕ_2) should be similar in complexity to $r(z)$, I only use ϕ_1 and ϕ_2 that are also linear in logs.

$$\begin{aligned}\Theta_{\text{CD,LPA}} = \{\theta : & \theta_z \geq 0, q - \theta_1 - z^\top \theta_z = a, a \geq 0, \\ & \operatorname{cov}[a\phi_1(q,k), \phi_2(q,k)] \geq 0, \\ & \phi_1, \phi_2 \in \{\phi : \phi = \varphi_1 + \varphi_k k + \varphi_q q, \varphi_k \geq 0, \varphi_q \geq 0, \varphi_1 + \varphi_k k + \varphi_q q \geq 0\}\}\end{aligned} \tag{39}$$

The set $\Theta_{\text{CD,LPA}}$ can be constructed without considering infinitely many values of φ. Theorem 6 shows we only need to consider a finite set of values of φ so the identified set is a finite number of linear inequalities.

[17] If we applied the full restrictions, the bounds on τ would be the values of two semi-infinite linear programming problem. Semi-infinite linear programming problems can be solved by a method known as discretization. Discretization solves the problem by solving a sequence of finite linear programming problems which each use only a finite number of the constraints. The method increases the number of constraints used until the value of the linear programming problems converges (see Still 2001). So one way to solve the above problems and compute bounds on τ is to choose a sequence of increasing functions for (ϕ_1, ϕ_2) that eventually contains all relevant increasing functions and solve the sequence of finite problems that use only the first $N, N+1, N+2, \ldots$ elements of the sequence. I show how to use a sequence of step functions to do this in Appendix A.

Theorem 6. Assume that the support of the observed distribution of (q, z, k) is a subset of the rectangle: $\left[\underline{q}, \overline{q}\right] \times [\underline{z}_1, \overline{z}_1] \times \cdots \times [\underline{z}_L, \overline{z}_L] \times \left[\underline{k}, \overline{k}\right]$. $\Theta_{\text{CD,LPA}}$ can be equivalently expressed as,

$$\begin{aligned}\Theta_{\text{CD,LPA}} = \{\theta :& \theta_z \geq 0, q - \theta_1 - z^\top \theta_z = a, \\ & \underline{q} - \theta_1 - \overline{z}^\top \theta_z \geq 0, \\ & \text{cov}\left[a\phi_1(q,k), \phi_2(q,k)\right] \geq 0, \\ & \phi_1, \phi_2 \in \left\{\phi : \phi = (k - \underline{k}), \phi = \left(q - \underline{q}\right), \text{ or } \phi = 1\right\}\Big\}\end{aligned} \tag{40}$$

Proof. See Appendix B. □

The bounds on τ can then be computed by solving two finite linear programming problems, a trivial computation,

$$\text{max or min }_{\theta \in \Theta_{\text{CD,LPA}}} \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t \left(1, z_t^\top\right)^\top\right] \theta. \tag{41}$$

3.3 Bayesian inference using results from Kline and Tamer (2016)

But we do not know the parameters of the linear programming problems that define the bounds on τ. We have to estimate them from data. So we need to account for estimation error in constructing the bounds.

The bounds on τ are *deterministic* functions of a finite vector of moments which I will call μ. The parameters μ are expectations of random variables which determine the covariance restrictions in $\Theta_{\text{CD,LPA}}$ and the moments in the objective function: $\mathbb{E}\left[vx^\top\right]$, $\mathbb{E}[vy]$, and $\mathbb{E}\left[v\left(1, z^\top\right)^\top\right]$. Let $\tau_{LB}(\mu)$ and $\tau_{UB}(\mu)$ be the lower and upper bound on τ for a given vector of reduced-form parameters μ.

I use a Bayesian method from Kline and Tamer (2016) to make inference on the identified set for τ, $[\tau_{LB}, \tau_{UB}]$. Let $\pi(\mu|\text{data})$ be the posterior distribution of μ given the data. Let $\Phi_n(\mu)$ be the multivariate normal distribution with mean $\widehat{\mu}$ and variance matrix $n^{-1}\Sigma$ where $\widehat{\mu}$ is the maximum likelihood estimator of μ and Σ is the variance-covariance matrix of the estimate. In practice, I use the sample means for $\widehat{\mu}$ and the sample covariance matrix for Σ.

Because the vector μ is a reduced-form parameter, Bernstein-von Mises theorems imply the posterior distribution of μ converges to a normal distribution centered around the true parameter value, see Bickel and Kleijn (2012) and the results they cite. These theorems imply that $\pi(\mu|\text{data})$ and $\Phi_n(\mu)$ converge in total variation as sample size increases,

$$\sup_h \left| \int h d\pi(\mu|\text{data}) - \int h d\Phi_n(\mu) \right| \to 0, \tag{42}$$

independently of the prior specified for μ[18]. This result allows us to use the distribution $\Phi_n(\mu)$ as an approximation to the posterior distribution $\pi(\mu|\text{data})$ in large samples without specifying a prior for μ directly.

For sufficiently large samples, the posterior probability that a given value τ_0 is in the identified set is,

$$\rho_n(\tau_0) = \int_\mu \mathbf{1}\left\{\tau_0 \in [\tau_{LB}(\mu), \tau_{UB}(\mu)]\right\} d\Phi_n(\mu), \tag{43}$$

where the posterior of μ is approximated by Φ_n. ρ_n is our posterior belief about whether τ_0 belongs to the set identified by the linear positive association assumptions. It does not require computing posterior beliefs about τ itself. Posterior beliefs about τ would always depend on our prior no matter how much data we accumulate because τ is only partially identified by the data. So it is more convenient to make inference on the bounds directly because the bounds are point-identified.

I use the $\rho_n(\cdot)$ function to form what I call a "Bayesian hypothesis interval" (BHI). Frequentist confidence intervals can be constructed by searching for parameter values that are not rejected by the α hypothesis test. We can construct a similar interval from a Bayesian perspective by finding the set of parameter values τ_0 such that our posterior belief about the likelihood a certain parameter value belongs to the identified set exceeds α. I call the resulting interval, a "Bayesian hypothesis interval", because it mimics the hypothesis test characterization of the frequentist confidence interval.

I use hypothesis intervals instead of Bayesian credible intervals for a couple of reasons. First, hypothesis intervals are uniquely defined while credible intervals are any interval that is more likely than a certain cut-off probability. There are many valid credible intervals of a given size. More critically, credible intervals that cover τ depend on the prior specified for τ even in very large samples. τ is only partially identified so the posterior distribution does not concentrate around a single point. Hypothesis intervals only depend on the posterior distribution for the identified parameters τ_{LB} and τ_{UB}.

The hypothesis interval is defined as,

$$HI_{\alpha,n} = \{\tau_0 : \rho_n(\tau_0) \geq \alpha\}. \tag{44}$$

So the 10% hypothesis interval is the set of parameter values (τ) that we believe, given the data, are more than 10% likely to satisfy the linear positive association assumption.

To compute the hypothesis interval in practice, do the following steps:

(1) Take B draws from the distribution $N\left(\widehat{\mu}, n^{-1}\widehat{\Sigma}\right)$: $\mu^{(1)}, \ldots, \mu^{(B)}$.

(2) For each $\mu^{(b)}$, compute: $\tau_{LB}^{(b)} = \tau_{LB}\left(\mu^{(b)}\right)$ and $\tau_{UB}^{(b)} = \tau_{UB}\left(\mu^{(b)}\right)$ by solving the two linear programs.

[18] So long as the prior's support includes the reals or otherwise includes a neighborhood of the (frequentist) probability limit of $\widehat{\mu}$

(3) The hypothesis interval includes all τ_0 such that:

$$\frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left(\tau_{LB}^{(b)} \leq \tau_0 \leq \tau_{UB}^{(b)}\right) \geq \alpha \tag{45}$$

3.4 Why Bayesian inference?

Bayesian inference has practical advantages over frequentist inference in partially identified models. When the bounds on the parameter of interest are non-differentiable but known functions of point-identified parameters, Bayesian methods can simply sample from the posterior distribution for the point-identified parameters and apply the functions to recover posterior beliefs about the parameter of interest. $\tau_{LB}(\mu)$ and $\tau_{UB}(\mu)$ are examples of such functions. The non-smoothness of $\tau_{LB}(\mu)$ and $\tau_{UB}(\mu)$ poses no difficulty for Bayesian inference because the posterior is conditional on the data. We do not need to consider how the distribution of the estimated parameters μ changes as sample size increases. But, to form the asymptotic distribution for frequentist inference, we would have to consider exactly that which complicates frequentist inference on non-smooth functions of means[19].

In general, frequentist and Bayesian inference will not coincide in partially identified models (see Moon and Schorfheide 2012), but see Kline and Tamer (2016) for when they do in large samples.

4 Why not use a proxy or instrumental variable approach?

Having now fully developed this new partial identification approach, I now answer the question, "Why?" Why do I not pursue a modification of the proxy approach or use a direct instrumental variables to resolve the identification problem?

First, I examine the proxy approach and study what it rules out relative to the model I developed. The following transition equation is the proxy model's main identification assumption as used in Doraszelski and Jaumandreu (2013) and De Loecker (2013),

$$a_t = g\left(a_{t-1}, x_{t-1}\right) + \eta_t, \quad \mathbb{E}\left[\eta_t | a_{t-1}, x_{t-1}, k_t\right] = 0 \tag{46}$$

[19] Hsieh, Shi, and Shum (2017) propose one frequentist method of making inference on the value of a linear program based on performing hypothesis tests on the Kuhn-Tucker optimality conditions. Similar to most methods for frequentist inference in partially identified models (see Chernozhukov, Hong, and Tamer 2007 and Chernozhukov, Lee, and Rosen 2013 for other examples studying related problems), we would need to choose tuning parameters to select potentially binding moments. We need to do this to have reasonable power if we have a decent number of constraints or else the number of moments is equal to the dimension of the Kuhn Tucker conditions, which is large. We would likely need to use something similar to the generalized moment selection procedure of Andrews and Soares (2010). There is limited theory available on how best to choose these tuning parameters in this literature.

where η_t is a shock to productivity and x_{t-1} are observed variables representing the endogenous investment plants make in their productivity. The identification strategy in the proxy literature is based on making assumptions about information sets: assumptions about *what* is known *when*. The proxy approach assumes that while η_t is known in period t, it is unknown in period $(t-1)$. It assumes the choices (x_{t-1}, k_t) are made in period $(t-1)$ so they can not be made on the basis of η_t, justifying the above moment restrictions.

This structure rules out much of what the general model of productivity choice I propose allows because η_t as defined above will either be partially known in $(t-1)$ or because x_{t-1} will be chosen partially in period t. For example, the following are ruled out by the proxy model, but allowed in my general framework[20]:

- **Flexible Productivity.** Productivity is flexible if it can be adjusted in the current period. This is ruled out by the proxy model because either (1) x_{t-1} controls for this flexible choice, in which case it is correlated with η_t because x_{t-1} then must be in the period t information set or (2) x_{t-1} does not control for this choice and η_t is chosen flexibly by the plant in response to the plant's current state variables, among which are (a_{t-1}, x_{t-1}, k_t) so η_t is not uncorrelated with them as the moment condition requires. An example of flexible productivity is workers putting in greater effort because a firm introduced a new performance incentive.

- **Unobserved Adjustments to Productivity.** Even if productivity is inflexible and a firm makes all its decisions about productivity in period $t-1$, if some of that investment is not included in x_{t-1}, the proxy model will not be identified. If x_{t-1} does not completely control for prior investment in productivity, some of that investment is included in η_t and that investment was made with knowledge of either the $(t-1)$ information set (inflexible productivity) or the t information set (flexible productivity). In either case, η_t will be correlated with variables in the $(t-1)$ information set, like (a_{t-1}, x_{t-1}, k_t).

  This is important because it may seem like when we want to understand the causal effect of something on productivity we can just include the variable of interest in x_{t-1} like De Loecker (2013) does for exporter status. He wants to find the causal effect of exporting on productivity, not just the selection effect described in Melitz (2003). It is very reasonable to assume that entry into exporter status in period t is uncorrelated with information that only became available in period t. But if we think plants can learn from exporting, we should also think plants can learn in other ways or, more generally, that plants have some control over their productivity. Unless exporter status *fully* explains plant investment decisions in productivity, we are not identified.

Direct instrumental variable approaches also prove difficult in this setting. I will put to the side the important data issues with finding instruments in common production datasets and ask, "What would the ideal instrument look like?" Let's write the production equation out

[20] Here I focus on the differences in the model of productivity not on the other generalizations I make relative to the proxy model. For example, I also allow for multiple unobserved state variables that cannot be proxied by observables and capacity constraints which are ruled out by the proxy model.

using the notation from the model in Section 2,

$$q_t = f\left(z_t\left(s\right), k_t\left(s\right)\right) + a_t\left(s\right) \tag{47}$$

(z, k) are determined by exactly the same state variables as a. The only difference between productivity and inputs is that we observe inputs and do not observe productivity. Therefore, finding a variable that is correlated with one set of factors of production (observed inputs) but not correlated with the choice of another factor of production (productivity) is difficult without introducing restrictions that make productivity in some way different than the other factors of production. For example, a demand shock (a change in ξ) would naturally affect the choice of *all* factors of production, observed inputs and productivity. A shock to the price of an observed input would cause substitution to or from productivity like it would for all other factors of production.

Of course, we could assume that productivity cannot react to current period shocks to demand or input prices, but that observed inputs can, a timing restriction that effectively allows some state variables to affect input choice but not productivity. But such a restriction is difficult to motivate when we are not sure exactly what productivity is. Is productivity like capital or capacity (fixed in the current period) or is it effort (which can be flexibly adjusted)? The goal of this paper is to avoid such timing assumptions and discover what we can learn from a general economic model of productivity choice.

5 How restructuring in the electricity industry affected productivity

Having developed a method for bounding the coefficients of productivity regressions, I now study whether the bounds are narrow enough for us to learn anything about real empirical problems. Does linear positive association and the data restrict the underlying parameters of interest enough to update our priors? I answer this question by studying how restructuring in the electricity industry affected power plant productivity using the approach developed in this article.

5.1 Industry and policy background

I start by briefly describing the history of regulation in the electricity industry and how the restructuring policy movement changed how electricity generators were regulated in some US states.

Historically, electric utilities were vertically integrated across the three stages of production: generation, transmission, and retail to end consumers. The price utilities received for the electricity they produced was tied to the costs they incurred. Roughly, regulations set price to the utility's average cost of producing the electricity plus a regulated rate of return on capital investments[21]. If a utility lowered its costs, it would receive a lower price, giving

[21] In some states, rewards were not based on average cost but on some approximation of marginal cost,

it less of an incentive to reduce costs and invest in higher productivity. In the mid-to-late 1990's, several US states restructured their electricity markets. They split up the vertically-integrated utilities and let competition among the firms generating electricity determine the price of electricity. By disintegrating the utilities, they were able to regulate the naturally monopolistic stage of electricity production, electricity transmission, while allowing for competition in the generation stage. I study what effect introducing price competition in the generation stage had on power plant productivity. See Borenstein and Bushnell (2015) for a fuller history of the restructuring policy.

Restructuring gives power plants an incentive to be more productive because, if they have lower costs than their competitors, they can earn greater markups. But it also has several effects that might reduce productivity. First, there may have been efficiencies to being integrated across generation and transmission. Second, to the extent productivity is a form of unobserved capital, Averch and Johnson (1962) argued rate of return regulation can cause over-investment in capital inputs. If there was such an over-investment and restructuring caused a reduction in productivity as plants drew back those investments, it may be associated with an *increase* in welfare (see Section 5.8). Lastly, restructuring increased the level of competition in the industry both by introducing price competition and by making it easier for power producers not owned by electric utilities (independent power producers) to enter the market. Increases in competition do not necessarily encourage firms to be more productive, see Vives (2008), the Schumpeterian growth literature (Aghion and Howitt, 1992), and patents. Greater competition reduces future profits so plants see less advantage to making investments in productivity.

I study how restructuring changed power plant productivity in practice. I am not the first to do so, see Fabrizio, Rose, and Wolfram (2007) and Knittel (2002). My contribution to this literature is that I allow power plants to adjust their productivity in response to the incentives offered by restructuring. I allow productivity to be a choice.

5.1.1 How might power plants choose productivity?

Is there any direct evidence that power plants and their employees can make choices that affect productivity? Bushnell and Wolfram (2009) find that the identity of the power plant *operator* affected the power plant's heat rate (the ratio of fuel used to power produced). That this is possible suggests that there are actions operators can take that increase the efficiency of the power plant. Additionally, in interviews with people in the industry they found that people believe the "skill and effort" of employees makes a difference in plant performance.

> At first glance, it might seem that workers should have little scope to influence the performance of the electricity industry and that this should be particularly true of the generation sector, where costs are dominated by the capital required to build plants and the fuel required to operate them. Overall, labor costs constitute a small fraction of generation costs. Yet in extensive interviews with plant managers and utility executives in the United States and Europe, most

but the main point is that the price was not set by market forces. See Knittel (2002) for more on these other policies.

expressed the belief that the individual skill and effort of key personnel could make a significant difference in the performance of generating plants.

Bushnell and Wolfram (2009)

5.2 Data

I use the same data as Fabrizio, Rose, and Wolfram (2007) for the years 1981 to 1999[22]. The original data is from the Energy Information Administration (EIA), the Federal Energy Regulatory Commission (FERC), and the Rural Utilities Service (RUS).

The data is at the power plant-level and includes large (greater than 100 megawatt capacity), fossil fuel power plants from the years 1981 to 1999. The first restructuring policy changes were in 1996 so the data has information on power plants before and after a state restructured its electricity market.

I then extend the dataset from 2000 to 2003 using data from the EIA and FERC to capture additional information on the effect of the restructuring policy. It might have taken some time for utilities to adjust to the restructuring policy, for new plants to enter, and for capacity to adjust so including additional years should be helpful in identifying the full effect of restructuring. It takes about 3 years to build a natural gas power plant[23] and longer to build a coal plant.

I measure output by millions of megawatt-hours of electricity generated net of electricity the power plant uses itself, fuel use by units of heat energy (in millions of million British thermal units), capacity by the total nameplate capacity of the plant (in megawatts), labor by total employment, and non-fuel expenditures in millions of dollars. The data is annual.

I include only power plants with positive employment and non-fuel expenses. Additionally, some fuel data is clearly incorrect. Some plants have impossibly large or small fuel-to-output ratios, likely because the utility filled out the form using different units than the form requested. To deal with the measurement error, I use the average heat rates for different types of fossil fuel power plants published by the Energy Information Administration for the year 2004 and remove observations that have a heat rate more than twice the average heat rate or less than one-half the average rate.

There are two types of power plants in my dataset: investor-owned power plants and municipally-owned power plants. Investor-owned power plants were subject to cost-of-service regulation prior to restructuring and after restructuring competed in a market to sell the electricity they produced. Municipally-owned power plants were not subject to cost-of-service regulation by the public service commission but were controlled directly by the municipality. Restructuring did not change their regulatory situation.

Figure 1 shows the joint distribution of fuel and power and which observations I mark as mismeasured.

[22] I thank Fabrizio, Rose, and Wolfram very much for making their data available and easy to use.
[23] See NEA (2016).

Figure 1: Relationship between fuel and electricity generated (with mismeasured observations), observations between the two lines make up the dataset I use

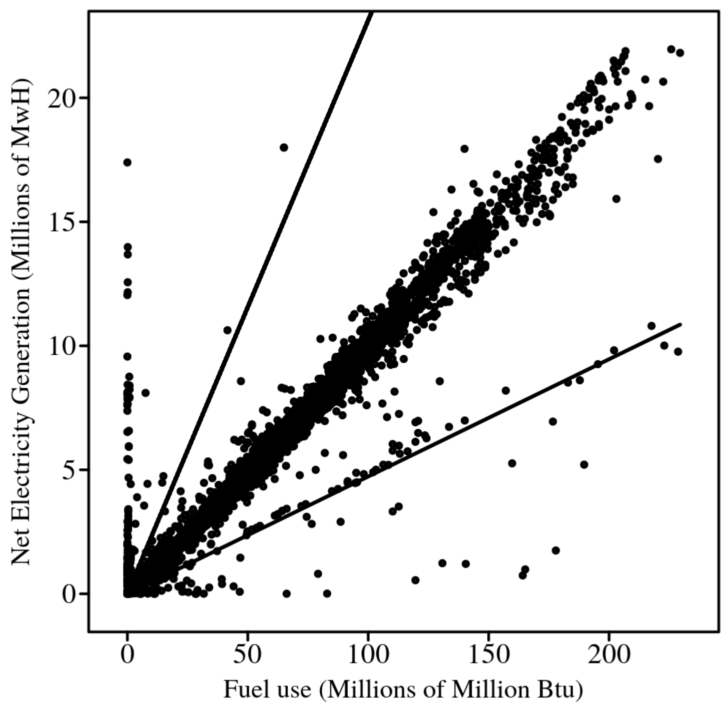

Table 1 gives descriptive statistics on the data with the mismeasured observations removed.

5.3 Empirical model of production

I build a model of electricity production to inform the shape of the production function I use empirically.

Let H be fuel burned by the power plant in units of heat energy, L be number of workers employed, and E be nonfuel expenditures (mostly maintenance and other operating costs).

Within a short window of time, a power plant can only produce more electricity by burning more fuel. Output increases roughly in proportion to the energy content of the fuel burned (H). Let $Q^{(\sigma)}$ be the power produced within a short window of time σ, $H^{(\sigma)}$ be the amount of fuel burned, and $K^{(\sigma)}$ be the available capacity within the short time window σ. I formalize the above description of electricity production into the following production equation,

$$Q^{(\sigma)} = \min \left\{ H^{(\sigma)} F\left(L, E\right) A, K^{(\sigma)} \right\}. \tag{48}$$

Our data is at the annual level, not in these short windows of time. I sum across these short time periods for the entire year to get the production equation I take to the data,

$$\sum Q^{(s)} = F\left(L, E\right) \sum_{s} H^{(s)} A \implies Q = F\left(L, E\right) HA, \tag{49}$$

Table 1: Descriptive statistics of data (with mismeasured observations removed)

Statistic	Plants	Output	Fuel	Capacity	Labor	Non-fuel
Mean	All	3.38	34.93	798	153	14.33
	IO	3.48	35.90	827	153	14.72
Median	All	2.10	22.15	575	112	9.86
	IO	2.22	23.28	614	114	10.20
Standard Deviation	All	3.69	37.15	666	133	14.43
	IO	3.75	37.71	678	131	14.83
Number of Plant-Years	$11,390$					
Range of years	1981 to 2003					

Notes. IO refers to power plants owned by investors. Output is in units of millions of megawatt hours, fuel is in units of millions of million British thermal units, capacity is in megawatts, labor is in number of employees, and non-fuel expenditures are in millions of dollars.

because $H^{(s)}F(L,E) \leq K^{(s)}$ for all time periods s.

I call A "fuel productivity" because it affects the productivity of burning fuel at the power plant. For the purposes of constructing the identified set, we can treat (Q/H) (fuel efficiency) as our output measure and (L,E) as our vector of inputs (Z from the theory in Section 2),

$$\frac{Q}{H} = F(L,E)\,A. \tag{50}$$

Although $Q = H \times F(L,E)\,A$, the output elasticity of H is *not* 1. Increasing H would cause more of the capacity constraints $K^{(s)}$ to bind throughout the year.

I use a Cobb Douglas functional form for F,

$$\frac{Q}{H} = L^{\theta_\ell} E^{\theta_e} A. \tag{51}$$

I assume $\theta \in \Theta_{\mathrm{CD,LPA}}$, the identified set constructed in Section 3.

5.3.1 Comparison of empirical model to Fabrizio, Rose, and Wolfram (2007)

In this brief section, I discuss the difference between this article's model of productivity and the model used in the most related empirical article (Fabrizio, Rose, and Wolfram, 2007). The measures of productivity used in the two articles are not directly comparable and should be interpreted slightly differently aside from the difference in identification strategies and the fact that I allow productivity to be endogenous. Fabrizio, Rose, and Wolfram (2007) regress log input on log output and other controls and policy variables to approximate a conditional input demand function and test how restructuring shifted the conditional input demand curve. My measure of productivity is a "multi factor productivity" measure where their productivity measure varies by the factor on the left hand side of the regression.

The most analogous results in Fabrizio, Rose, and Wolfram (2007) are their results for how restructuring affects the fuel conditional demand curve. They find that restructuring had a small, insignificant effect. I add a new result to this literature: the effect of restructuring on fuel productivity, allowing productivity to be endogenous.

5.4 Identification of the effect of restructuring

Restructured and non-restructured states differ aside from the restructuring policy itself. The average effect of restructuring is not the average difference in productivity across the two types of states. I identify the causal effect of restructuring on productivity and other outcomes by using a difference-in-difference-like strategy and plant fixed effects. I consider two different control groups to estimate the effect of restructuring.

5.4.1 Control Group 1: Investor-owned power plants in regulated states

My first identification strategy compares investor-owned power plants in restructured states to investor-owned power plants in regulated states, holding plant level productivity differences and aggregate productivity shocks fixed.

Suppose we want to estimate the causal effect of restructuring on some outcome X_{it}. Let EverRestruct_{it} be 1 if the state the plant is in ever restructured and zero otherwise and let Y_{it} be the years since restructuring or 0 if the state the plant is in never restructured (equivalent to interacting Y_{it} with EverRestruct_{it}). I estimate the following regression using data only on investor-owned power plants,

$$\begin{aligned} x_{it} = {} & \tau_X \times \text{Restruct}_{it} + \overline{\tau}_X Y_{it} \mathbf{1}\left(Y_{it} > 0\right) + \underline{\tau}_X Y_{it} + \kappa_X m_{rt} & (52) \\ & + \mu_{20} \times \mathbf{1}\left(\text{Age}_{it} \leq 20\right) + \mu_{30} \times \mathbf{1}\left(\text{Age}_{it} \leq 30\right) & (53) \\ & + \mu_{40} \times \mathbf{1}\left(\text{Age}_{it} \leq 40\right) + \iota \times \text{EverRestruct}_{it} + \xi_i + \delta_t + u_{it}, & (54) \end{aligned}$$

where i indexes plants, t indexes year, ξ_i is the plant-level fixed effect, δ_t is the year-level fixed effect, m_{rt} is a measure of market size (log population in the plant's Census division), and Restruct_{it} is 1 if the power plant is in a state where the restructuring law has passed and is 0 otherwise. The age of the power plant is the number of years since its construction. 20, 30, and 40 are roughly the 25-th, 50-th, and 75-th quantile of the age distribution.

The regression model above allows for a pre- and post-trend in restructured states to control for non-restructuring related restructured-state specific trends and also for potentially dynamic effects of restructuring. It uses plant-level fixed effects to control for permanent differences in productivity across plants unrelated to restructuring.

The inclusion of the plant level fixed effects is important for interpreting the estimated effect. τ_X is the average effect of restructuring on an *individual plant*, the within-plant effect. It is *not* the selection effect of the restructuring policy. Restructuring may make more productive plants more or less competitive leading to entry or exit, but that is not included in τ_X. A policy can have a selection effect without power plants having any control over their productivity. On the other hand, if we think power plants decide to be more or less productive in response to the new incentives offered by restructuring, then plants must be able to adjust their productivity. The plant-level effect of restructuring demonstrates the empirical relevance of allowing for productivity choice.

5.4.2 Control Group 2: Municipally-owned power plants in restructured states

The specification in Section 5.4.1 fails to identify the causal effect of restructuring if there is a shock unrelated to restructuring but specific to restructured states that occurred in the same year as restructuring. Such a shock would be confused with the restructuring policy itself and be included in the estimate of restructuring.

To deal with this concern, I also estimate the effect of restructuring by using municipally-owned power plants in restructured states as a control group. Because municipally-owned power plants were not directly affected by the policy, the effect of restructuring on municipally-owned power plants captures shocks in restructured states unrelated to the restructuring policy.

This specification uses both investor-owned and municipally-owned power plants:

$$\begin{aligned}
x_{it} = {} & \tau_X \times \text{Restruct}_{it}\text{IO}_{it} + \beta \times \text{Restruct}_{it} + \overline{\tau} Y_{it}\mathbf{1}\left(Y_{it} > 0\right)\text{IO}_{it} + \underline{\tau} Y_{it}\text{IO}_{it} && (55)\\
& + \overline{\beta} \times Y_{it}\mathbf{1}\left(Y_{it} > 0\right) + \underline{\beta} \times Y_{it} + \kappa m_{rt} && (56)\\
& + \mu_{20} \times \mathbf{1}\left(\text{Age}_{it} \leq 20\right) + \mu_{30} \times \mathbf{1}\left(\text{Age}_{it} \leq 30\right) + \mu_{40} \times \mathbf{1}\left(\text{Age}_{it} \leq 40\right) && (57)\\
& + \iota \times \text{EverRestruct}_{it} + \iota_{IOU} \times \text{EverRestruct}_{it}\text{IO}_{it} + \xi_i + \delta_t + u_{it}, && (58)
\end{aligned}$$

where IO indicates whether the plant is investor-owned. This specification measures the causal effect of restructuring by how the outcome x changes within an investor owned plant relative to how it changes within a municipality owned plant when restructuring happens, controlling for pre and post trends, market size, and aggregate productivity shocks.

But there is a tradeoff. The incentives of the owners of municipally-owned power plants, municipalities, are different than the incentives of private investors and comparison between the two plant types could face its own issues. I use both approaches to check the robustness of my results.

5.5 How restructuring affected fuel efficiency, output, and input use

I start my empirical analysis by estimating the effect of restructuring on fuel efficiency (Q/H). Much like labor productivity (Q/L), fuel efficiency is influenced by other inputs in the plant's production function aside from productivity. But unlike structural productivity measures, fuel efficiency can be simply read from the data. Because it does not have to be identified, it is a useful baseline.

I find the effect of restructuring on log fuel efficiency (τ_{q-h}) is *negative*. $\tau_{q-h} = -1.32\%$ if we use regulated investor-owned power plants as the control group or $\tau_{q-h} = -2.92\%$ if we use municipally-owned power plants in restructured states as the control group.

But I also find that restructuring caused power plants to use fewer nonfuel inputs (see Table 2). Plant output fell by 10% after restructuring. Because the use of other inputs increases fuel efficiency, one explanation for the fall in fuel efficiency is simply that power plants used fewer other factors of production as a result of restructuring. We cannot tell from the fact that fuel *efficiency* fell whether fuel *productivity* decreased.

I use the partial identification method proposed in the previous sections of this article to determine whether decreases in productivity were responsible for decreases in fuel efficiency or whether the drop can be explained entirely by lower factor use.

5.6 The effect of restructuring on power plant productivity

The effect of restructuring on fuel productivity is,

$$\tau_a = \tau_{q-h} - \tau_{f(\ell,e)}, \tag{59}$$

where τ_a is the effect of restructuring on log fuel productivity, τ_{q-h} is the effect of restructuring on log fuel efficiency (Q/H), and $\tau_{f(\ell,e)}$ is the effect of restructuring on the log production function. τ_{q-h} is point-identified in the data because $(q-h)$ is observed, but $\tau_{f(\cdot)}$ is not because the log production function $f(\ell, e)$ is unknown.

Because $\tau_{\alpha X+b} = \alpha\tau_X$ for constants α and b, using the Cobb Douglas specification for f we can write,

$$\tau_f = \theta_\ell\tau_\ell + \theta_e\tau_e, \tag{60}$$

where τ_ℓ and τ_e are estimates of the effect of restructuring on log labor and log nonfuel expenditures. So the effect of restructuring on productivity is a weighted sum of the effects of restructuring on fuel efficiency and input use. I estimated τ_{q-h}, τ_ℓ, and τ_e in Section 5.5 so the only unknown parameter left is θ. I assume the unknown parameter θ belongs to the set $\Theta_{\text{LPA,CD}}$ introduced in Section 3 and construct bounds on τ_a given this assumption.

To make the identification assumptions less of a "black box", I work through how they pin down τ_a. Because $(\theta_\ell, \theta_e) \geq 0$, if restructuring has a negative effect on input use ($(\tau_\ell, \tau_e) \leq 0$) as we saw in Section 5.5, then restructuring will have a negative effect on the value of the log production function f ($\tau_f = \theta_\ell\tau_\ell + \theta_e\tau_e \leq 0$). So:

$$\tau_a = \tau_{q-h} - \tau_f \geq \tau_{q-h}. \tag{61}$$

Because $\tau_{q-h} < 0$ (see Section 5.5), the question is whether (θ_ℓ, θ_e) can be large enough that $\tau_a = \tau_{q-h} - \tau_\ell\theta_\ell - \tau_e\theta_e$ can be positive or whether the linear positive association assumption

Table 2: Effect of restructuring on plant fuel efficiency and plant factor use

Dependent Variable	Control Group	τ	$\overline{\tau}$	$\underline{\tau}$	κ
Fuel Efficiency	IO	-1.32%	0.66%	-0.04%	2.67%
		(0.53%)	(0.20%)	(0.03%)	(0.75%)
	MUNI	-2.92%	-1.24%	0.09%	3.16%
		(1.19%)	(0.80%)	(0.07%)	(0.75%)
Power	IO	-10.01%	-2.69%	-1.66%	6.24%
		(3.31%)	(1.28%)	(0.21%)	(4.73%)
	MUNI	-6.04%	-19.56%	0.35%	8.16%
		(7.29%)	(4.90%)	(0.44%)	(4.57%)
Fuel	IO	-8.69%	-3.35%	-1.61%	3.57%
		(3.15%)	(1.21%)	(0.20%)	(4.50%)
	MUNI	-3.12%	-18.32%	0.26%	5.00%
		(6.91%)	(4.64%)	(0.42%)	(4.33%)
Capacity	IO	-2.48%	-1.23%	-0.31%	-2.87%
		(0.98%)	(0.38%)	(0.06%)	(1.40%)
	MUNI	-5.17%	-7.30%	-0.37%	-2.80%
		(2.08%)	(1.40%)	(0.13%)	(1.30%)
Labor	IO	-1.72%	1.38%	-0.95%	7.94%
		(1.57%)	(0.60%)	(0.10%)	(2.24%)
	MUNI	7.74%	-4.20%	-0.69%	9.06%
		(3.58%)	(2.41%)	(0.22%)	(2.25%)
Nonfuel Expenditures	IO	-3.84%	-3.65%	-0.26%	4.11%
		(2.29%)	(0.88%)	(0.15%)	(3.27%)
	MUNI	-6.33%	-9.26%	-1.00%	6.37%
		(5.14%)	(3.45%)	(0.31%)	(3.22%)

sufficiently restricts how large θ can be so that $\tau_f = \theta_\ell \tau_\ell + \theta_e \tau_e$ is not *that* negative[24]. If θ cannot be so large, then $\tau_a < 0$. Otherwise, the sign of τ_a is ambiguous.

I find the effect of restructuring is to *reduce* fuel productivity. The effect of restructuring on fuel productivity is between -4.53% and -0.46% across the two control group specifications using the 10% hypothesis intervals. Point estimates of the bounds are between -2.87% and -1.12%. An effect magnitude between 0.46% and 4.53% is economically meaningful: to maintain the same output as before, the plant would need to burn 0.46% to 4.53% more fuel if it has the same labor and non-fuel expenditures.

We can quantify what a percentage increase in fuel use means. Fuel costs about \$3.25 per million British thermal units (EIA data). Multiplying that by the average fuel use in 2003 tells us that a 1% increase in fuel use costs $\$1,085,802$ in 2003 dollars. A 1% increase in fuel use would, by the same calculation, require about $17,730$ metric tons more carbon dioxide to be emitted at natural gas power plants. Taking the low-end of the EPA's Social Cost of Carbon (\$11 per metric ton of carbon dioxide), this costs society about $\$195,030$ for a natural gas power plant using an average amount of fuel.

The negative effect of restructuring could either be a result of the combination of the competitive effects of restructuring, the Averch and Johnson (1962) effect, and inefficiencies from disintegration dominating the incentive effects of restructuring, *or* it could be the effect of the policy was negatively initially because it *disrupted* the power plant's operations and, while the plant adjusted to the new regime, its productivity fell, see Holmes, Levine, and Schmitz (2012) for a theory of why disruptions can cause monopolists to innovate less than firms that face greater competition.

To investigate the second case, I study the slope of the post-restructuring trend ($\overline{\tau}$). We should be cautious in interpreting the trend as entirely causal because the risk of interpreting an exogenous investor-owned-power-plant-in-a-restructured-state trend as being a consequence of restructuring increases the further we move from the restructuring event itself. But the disruption story is worth considering as one explanation for the results. I estimate the annual trend in the effect of restructuring to be between 0.22% and 1.02% using investor-owned power plants as the control and between -0.84% and 0.96% using municipally-owned power plants as the control. Interpreting the post-trend entirely causally, the number of years until the policy effect is positive is (if $\overline{\tau} \geq 0$),

$$\eta = -1 \times \frac{\tau}{\overline{\tau}}. \tag{62}$$

Because the results using municipal plants as the control allow for $\overline{\tau} \leq 0$, the upper bound on the time it would take for the effect to be positive is unbounded. I study the lower bound on the time it takes for the sign of the restructuring effect to flip to investigate whether the disruption theory can be supported by the data. I form a lower bound on η by solving a fractional linear programming problem because both τ and $\overline{\tau}$ are linear functions

[24] If, alternatively, restructuring had predicted that input use would increase, then $\tau_f > 0$ and it would be unambiguous that $\tau_a < 0$.

of θ. Fractional linear programming problems can be transformed into linear programming problems so the same methods to bound the other parameters can be used to bound η[25].

The lower bound on η using Control Group 1 is 1.73 while the lower bound using Control Group 2 is 8.40. Because the maximum number of years post-restructuring in sample is 9, the results for Control Group 2 suggest that the productivity effect of restructuring is long-lasting. But the results using Control Group 1 suggest the sign may flip relatively soon. So the data might admit the disruption theory, but I do not have narrow enough bounds to conclude that it is the story across the two specifications. If Control Group 2 is the right specification, the results suggest there is no disruption effect, but if Control Group 1 is the right specification, the results suggest the disruption story plays a role. In either case, both specifications agree that the immediate effect of the policy is negative.

I also measure the effect of market size on productivity. Market size has a theoretically ambiguous effect. Larger markets both encourage entry, increasing competition and making the residual demand curve more elastic, *and* shift the demand curve outward. See the large literature on how market size affects market structure: Shaked and Sutton (1987), Sutton (1991), Bresnahan and Reiss (1991), Melitz (2003), and Syverson (2004)[26].

I find larger markets *increase* productivity choice at power plants. The elasticity of productivity with respect to market size is between 0.80% and 4.59%. The demand expansion effect of larger markets dominates the effect of encouraging entry. In the language of the model in Section 2, market size increases the ξ parameter at least on average. The result also provides evidence that demand shocks like market size affect productivity.

In fact, the result is difficult to explain with an exogenous productivity model. Although such models often find expansions in market size raise the minimum productivity required for entry (see Melitz 2003), the regression models all include plant level fixed effects so differences in entry thresholds across markets have limited ability to explain the coefficient. Intuitively, the coefficient means, "when market size is above its average over time within a market, plants are likely to have productivity above their own average" controlling for an aggregate time trend and plant age effects. It is the within-plant effect of an increase in market size.

Suppose we had two markets alike in every way except that one market is of size 1 and the other market is of size 2. Suppose that productivity is exogenous and firms make a standard entry decision. If the model is Melitz-like, the average productivity of the surviving firms in the market of size 1 will be less than the average productivity of the surviving firms in the market of size 2. But market size will be uncorrelated with productivity conditional on the fixed characteristics of the plants that survive in both markets (i.e. if we include plant fixed effects) because, in an exogenous productivity model, identical plants will have identical

[25] The only requirement to transform a fractional linear program into a linear program is that the denominator must be constrained in the programming problem to be positive. Because I am only computing the lower bound for η, it is fine to add a constraint like $\overline{\tau} \geq 0.0001$ because this constraint will not effect the solution to the problem. The only effect of adding the constraint is to allow me to reformulate the fractional linear program as a linear program.

[26] The two Sutton articles on "endogenous sunk costs" are especially related to the current problem because endogenous productivity can be viewed as such a cost.

Table 3: Effect of restructuring on plant fuel productivity

Parameter	Control Group	LB (10%)	LB	UB	UB (10%)
τ_a	MUNI	−4.53%	−2.87%	−2.44%	−0.98%
	IO	−2.01%	−1.44%	−1.12%	−0.46%
κ_a	MUNI	0.88%	2.33%	3.09%	4.38%
	IO	0.99%	2.31%	2.95%	4.20%
$\overline{\tau}_a$	MUNI	−0.84%	−0.22%	0.32%	0.96%
	IO	0.22%	0.54%	0.68%	1.02%
$\underline{\tau}_a$	MUNI	0.09%	0.22%	0.36%	0.51%
	IO	−0.08%	−0.04%	0.08%	0.12%
β	MUNI	−0.54%	0.79%	1.45%	3.19%
$\overline{\beta}$	MUNI	−0.22%	0.28%	0.77%	1.28%
$\underline{\beta}$	MUNI	−0.46%	−0.33%	−0.25%	−0.13%

Notes. Columns "LB (X%)" and "UB (X%)" give the upper and lower limits of the $X\%$ credible interval for the parameter (containing all parameter estimates that are at least $X\%$ likely to belong to the identified set as measured by the posterior distribution). LB and UB refer to the "point estimates" of the bounds. IO refers to investor-owned power plants in non-restructured states (control group 1). MUNI refers to municipally-owned power plants in restructured states (control group 2).

productivities in both markets. In an endogenous productivity model, identical plants may have *different* productivities in each market so including plant fixed effects would not make the coefficient on market size zero.

To the extent that the controls included in my regression model do not make markets "identical in every way" except market size or control for all plant characteristics, exogenous productivity models still may be able to explain the coefficient, but the non-zero coefficient on market size given the controls I do have is at least suggestive of productivity choice.

5.7 The effect of restructuring on aggregate productivity

I also study how restructuring affected output share-weighted state-level productivity ("aggregate productivity"). The effect of restructuring on aggregate productivity may differ from the plant-level effect because the policy may change the correlation between productivity and output share. Aggregate productivity will increase both when average plant productivity increases and when output is reallocated across power plants increasing the correlation between productivity and market share. Since at least Olley and Pakes (1996), a main concern of the productivity literature has been estimating how different policies affect the relationship between market share and productivity. I also study this relationship and whether restructuring reallocated output towards more productive power plants.

I call the aggregate productivity measure I use "geometric aggregate productivity" (GAP) to differentiate it from the measure used in Olley and Pakes (1996) because it is the geometric average of productivity weighted by output-share (Olley and Pakes 1996 use the arithmetic mean). The only reason I use the geometric mean is that forming the bounds is computationally simpler because the bounds are the solution to linear programs following Section 3,

$$o_{it} = \frac{Q_{it}}{\sum_{j \in \text{state}} Q_{jt}} \tag{63}$$

$$\text{GAP}_{\text{state},t} = \prod_{i \in \text{state}} A_{it}^{o_{it}}. \tag{64}$$

Following Olley and Pakes (1996), I decompose aggregate productivity into an average productivity term and a reallocation term which measures the covariance between market share and productivity:

$$\log \text{GAP} = \underbrace{\frac{1}{N}\sum_i a_i}_{\text{Average Productivity}} + \underbrace{\sum_i \left(a_i - \frac{1}{N}\sum_i a_i\right) \times \left(o_i - \frac{1}{N}\sum_i o_i\right)}_{\text{Reallocation term}}. \tag{65}$$

I estimate the effect of restructuring on the two terms both separately and in combination. I use a regression model in the same vein as the specifications I used for the plant-level effects,

$$x_{\text{state},t} = \tau_x \times \text{Restruct}_{\text{state},t} + \underline{\tau} Y_{\text{state},t} + \overline{\tau} Y_{\text{state},t} \times \mathbf{1}\left(Y_{\text{state},t} > 0\right) + \text{Year fixed effect} \tag{66}$$

$$+ \text{State effect} + u_{\text{state},t}, \tag{67}$$

where x is either log GAP, average log productivity, or the reallocation term. I estimate the regression both using only investor-owned power plants to compute aggregate productivity and using all power plants.

Restructuring caused average log productivity in a state to fall between -5.50% and -1.88% for investor-owned utilities and by -5.59% and -0.92% for all power plants. Average state productivity fell like plant-level productivity fell. But among investor-owned power plants, the plants directly affected by the policy, the reallocation effect of the policy was *positive*, between 0.77% and 3.09%. This result suggests that restructuring improved the allocation of output so that it went towards more productive power plants because output was now allocated across these plants by a market. The reallocation effect including all power plants has a positive point estimate for the lower bound, but the 10% hypothesis interval extends from -0.17% to 2.52%. The positive sign of the reallocation effect gives evidence that the new market for electricity more efficiently allocated output across the investor-owned power plants even though it reduced average productivity.

But the net effect of the reallocation and average productivity is negative. The effect of restructuring on GAP for all power plants is between -4.02% and -0.21% (point estimate

Table 4: Effect of restructuring on geometric aggregate fuel productivity (GAP)

Variable	Parameter	Plants Included	LB (10%)	LB	UB	UB (10%)
GAP	τ	IO	−3.64%	−2.17%	−1.44%	0.26%
		All	−4.02%	−2.29%	−1.78%	−0.21%
	$\overline{\tau}$	IO	−0.24%	0.31%	0.91%	1.49%
		All	0.08%	0.70%	1.04%	1.59%
	$\underline{\tau}$	IO	−0.14%	−0.03%	0.06%	0.17%
		All	−0.11%	−0.03%	0.09%	0.19%
Average Productivity	τ	IO	−5.50%	−3.93%	−3.43f%	−1.88%
		All	−5.59%	−3.81%	−2.69%	−0.92%
	$\overline{\tau}$	IO	−0.45%	0.08%	1.08%	1.64%
		All	−0.23%	0.34%	0.75%	1.28%
	$\underline{\tau}$	IO	−0.00%	0.10%	0.22%	0.32%
		All	−0.01%	0.09%	0.20%	0.30%
Reallocation	τ	IO	0.77%	1.76%	2.14%	3.09%
		All	−0.17%	0.85%	1.53%	2.52%
	$\overline{\tau}$	IO	−0.68%	−0.38%	0.23%	0.50%
		All	−0.29%	0.06%	0.36%	0.69%
	$\underline{\tau}$	IO	−0.23%	−0.16%	−0.13%	−0.07%
		All	−0.18%	−0.12%	−0.09%	−0.03%

Notes. Notes. Columns "LB (X%)" and "UB (X%)" give the upper and lower limits of the X% credible interval for the parameter (containing all parameter estimates that are at least X% likely to belong to the identified set as measured by the posterior distribution). LB and UB refer to the "point estimates" of the bounds. IO results look at aggregate productivity for investor-owned power plants only.

between −2.29% and −1.78%) and among only investor-owned power plants GAP is between −3.64% and 0.26% (point estimate between −2.17% and −1.44%).

5.8 Welfare interpretations

These results are not a welfare analysis of the restructuring policy. The negative fuel productivity effect of restructuring by itself is not evidence that restructuring had a negative effect on total surplus. It may be efficient for power plants to reduce their productivity because, in the model of Section 2, productivity is a factor of production bought at a cost so plants being more productive is not an unambiguous social benefit.

In fact, because rate-of-return regulation rewards utilities with a regulated rate-of-return for a given capital investment, Averch and Johnson (1962) show utilities subject to rate-of-return regulation will over-invest in capital inputs, distorting the ratio of the marginal product of non-capital inputs to capital inputs from their efficient levels.

> If the rate of return allowed by the regulatory agency is greater than the cost of capital but is less than the rate of return that would be enjoyed by the firm were it free to maximize profit without regulatory constraint, then the firm will substitute capital for the other factor of production...
>
> Averch and Johnson (1962)

To the extent productivity is an unobserved capital investment, it may be the case that plants were investing inefficiently large amounts in fuel productivity and the reduction in fuel productivity after restructuring was *efficient*. In addition, fuel productivity is not the only outcome affected by restructuring. Restructuring also changed the electricity prices paid by consumers. The price effect of restructuring matters substantially for determining the total welfare effect of restructuring. The physical fuel productivity effect is only a part of the welfare effect.

6 Conclusion

Traditional production function estimation methodologies assume productivity is exogenous. I show that we can allow productivity to be a choice and meaningfully partially identify policy-relevant statistics of the productivity distribution. I prove a comparative static result that holds in a robust class of economic models and use it to construct a statistical restriction on the productivity distribution. I then propose a practical estimator for the identified set based on this restriction. The bounds are computed simply by solving two linear programming problems. Lastly, I apply the inference method to form bounds on the effect of restructuring in the electricity industry on power plant productivity.

References

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

Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica 60*(2), 323–351.

Andrews, D. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica 78*(1), 119–157.

Averch, H. and L. Johnson (1962). Behavior of the firm under regulatory constraint. *American Economic Review 52*(5), 1052–1069.

Bickel, P. and B. Kleijn (2012). The semiparametric bernstein-von mises theorem. *The Annals of Statistics 40*(1), 206–237.

Borenstein, S. and J. Bushnell (2015). The u.s. electricity industry after 20 years of restructuring. *NBER Working Paper* (21113).

Bresnahan, T. and P. Reiss (1991). Entry and competition in concentrated markets. *The Journal of Political Economy 99*(5), 977–1009.

Bugni, F., I. Canay, and X. Shi (2016). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*.

Bushnell, J. and C. Wolfram (2009). The guy at the controls: Labor quality and power plant efficiency. *International Differences in the Business Practices and Productivity of Firms* (3), 79–102.

Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica 75*(5), 1243–1284.

Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: estimation and inference. *Econometrica 81*(2), 667–737.

De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics 5*(3), 1–21.

De Loecker, J., J. Eeckhout, and G. Unger (2019). The rise of market power and the microeconomic implications. *Working paper*.

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

Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies 80*(4), 1338–1383.

Esary, J., F. Proschan, and D. Walkup (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics 38*(5), 1466–1474.

Fabrizio, K., N. Rose, and C. Wolfram (2007). Do markets reduce costs? assessing the impact of regulatory restructuring on us electric generation efficiency. *American Economic Review 97*(4), 1250–1277.

Flynn, Z., A. Gandhi, and J. Traina (2019). Measuring markups with production data. *Working paper*.

Gandhi, A., S. Navarro, and D. Rivers (2019). On the identification of production functions: how heterogeneous is productivity? *Working Paper*.

Griliches, Z. and D. Jorgenson (1967). The explanation of productivity change. *The Review of Economic Studies 34*(3), 249–283.

Griliches, Z. and J. Mairesse (1995). Production functions: the search for identification. *NBER Working Paper Series* (5067).

Holmes, T., D. Levine, and J. Schmitz (2012). Monopoly and the incentive to innovate when adoption involves switchover disruptions. *American Economic Journal: Microeconomics 4*(3), 1–33.

Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica 60*, 1127–1150.

Hsieh, Y.-W., X. Shi, and M. Shum (2017). Inference on estimators defined by mathematical programming. *Working Paper*.

Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica 50*(3), 649–670.

Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. *Working paper*.

Kline, B. and E. Tamer (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics 7*, 329–366.

Knittel, C. (2002). Alternative regulatory methods and firm efficiency: Stochastic frontier evidence from the u.s. electricity industry. *The Review of Economics and Statistics 84*(3), 530–540.

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

Manski, C. and J. Pepper (2000). Monotone instrumental variables, with an application to the returns to schooling. *Econometrica 68*(4), 997–1012.

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

Matzkin, R. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica 71*(5), 1339–1375.

Melitz, M. J. (2003). The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica 71*(6), 1695–1725.

Moon, H. R. and F. Schorfheide (2012). Bayesian and frequentist inference in partially identified models. *Econometrica 80*(2), 755–782.

NEA (2016). Nuclear energy agency press kits - economics of nuclear power faqs. https://www.oecd-nea.org/news/press-kits/economics-FAQ.html. Accessed: 2016-08-19.

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.

Shaked, A. and J. Sutton (1987). Product differentiation and industrial structure. *The Journal of Industrial Economics 36*(2), 131–146.

Still, G. (2001). Discretization in semi-infinite programming: the rate of approximation. *Mathematical Programming 91*(1), 53–69.

Sutton, J. (1991). *Sunk Costs and Market Structure.* MIT Press.

Syverson, C. (2004). Product substitutability and productivity dispersion. *The Review of Economics and Statistics 86*(2), 534–550.

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

Topkis, D. (1978). Minimizing a submodular function on a lattice. *Operations Research 26*(2), 305–321.

Van Biesebrock, J. (2003). The effect of technology choice on automobile assembly plant productivity. *The Review of Economic Studies 70*(1), 167–198.

Vives, X. (2008). Innovation and competitive pressure. *The Journal of Industrial Economics 56*(3), 419–469.

A Flexible estimation of the production function and identified set

In this appendix, I show how to construct the full linear positive association identified set to allow for bounds on highly flexible production functions. Instead of using only a subset of the increasing functions ϕ_1 and ϕ_2 that define the identified set $\mathcal{F}_{\text{LPA}}$, we will need to use all them.

There are many increasing functions, $\phi(q,k)$. How should we make sure we use all of them? Theorem 7 gives a useful result that simplifies constructing the set of all increasing functions used in the identified set $\mathcal{F}_{\text{LPA}}$. It shows that the identified set can be written equivalently as using all step functions so the set of functions to use can be indexed by a finite-dimensional vector of parameters.

Theorem 7. The following is an equivalent expression for $\mathcal{F}_{\text{LPA}}$,

$$\begin{aligned} \mathcal{F}_{\text{LPA}} = \{f : & Df \geq 0, q - f = a, a \geq 0, & (68) \\ & \operatorname{cov}\left[a\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq (u_{1,q}, u_{1,k})\right],\right. & (69) \\ & \left.\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq (u_{2,q}, u_{2,k})\right]\right] \geq 0 \quad \forall\left(u_{1,q}, u_{1,k}, u_{2,q}, u_{2,k}\right) \in [0,1]^4\Big\} & (70) \end{aligned}$$

where φ_q and φ_k are two strictly increasing functions from $\mathbb{R}$ to $[0,1]$.

Proof. See Appendix B. □

We can use this formulation to construct a flexible class of increasing functions to use to approximate the production function.

Suppose to approximate the production function, we use a vector of flexible basis functions indexed by a parameter m,

$$r_{p,m}(z,k) = \mathbf{1}\left(\left(\{\varphi_{z_\ell}(z_\ell)\}_{\ell=1}^{L}, \varphi_k(k)\right) \geq \left(\left\{\frac{p_{z_\ell}}{m}\right\}_{\ell=1}^{L}, \frac{p_k}{m}\right)\right), \tag{71}$$

Where p are restricted to be integers and all $p < m$ are included. Let $\theta_{p,m}$ be the coefficient on $r_{p,m}$. The requirement that the production function is increasing is equivalent to the requirement that $\theta_{p,m} \geq 0$ for $p \neq 0$ (write as $\theta_{-0} \geq 0$). The advantage of this flexible production function is that it is straightforward to control the complexity of the function and as $m \to \infty$ any increasing function will be approximated by the above basis. It would not be a good basis to use if we cared about bounding the derivative of the production function (because it is non-differentiable), but the only statistics of interest will be like τ which does not require a differentiable f.

We can then use an approximation to the identified set indexed by m_1 and m_2 both of which are no more flexible than m, $\max\{m_1, m_2\} \leq m$. The idea is to use only covariance restrictions that use functions (weakly) less flexible than the production function itself (using covariance restrictions much more flexible than the production function runs the risk of achieving narrow bounds via functional form).

$$\begin{aligned}
\Theta_{m_1,m_2,\text{LPA}} = \{\theta : &\theta_{-0} \geq 0, q - r(z,k)^\top \theta = a, a \geq 0, \\
&\text{cov}\left[a\mathbf{1}\left[(\varphi_q(q), \varphi_k(k)) \geq (u_{1,q}, u_{1,k})\right],\right. \\
&\left.\mathbf{1}\left[(\varphi_q(q), \varphi_k(k)) \geq (u_{2,q}, u_{2,k})\right]\right] \geq 0 \\
&\forall (u_{1,q}, u_{1,k}) \in \left\{0, \frac{1}{m_1}, \ldots, \frac{m_1 - 1}{m_1}\right\}^2 \text{ and} \\
&\left.\forall (u_{2,q}, u_{2,k}) \in \left\{0, \frac{1}{m_2} \ldots, \frac{m_2 - 1}{m_2}\right\}^2\right\}
\end{aligned} \tag{72}$$

Because $\Theta_{m_1,m_2,\text{LPA}}$ is a finite set of linear inequalities, the bounds on τ can be computed by solving two finite linear programming problems so computation is highly tractable and it is straightforward to construct the bounds given estimates of the unknown parameters of the linear programs. The linear programming problems are,

$$\text{max or min }_{\theta \in \Theta_{m_1,m_2,\text{LPA}}} \mathbb{E}\left[vx^\top\right]_{\text{left}}^{-1} \mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]_{\text{left}}^{-1} \mathbb{E}\left[v_t r(z_t, k_t)^\top\right] \theta. \tag{73}$$

Larger m will approach the nonparametric identified set, but of course, there will be a variance tradeoff in finite samples.

B Proofs

B.1 Proof of Lemma 1

Proof of Lemma 1.

From the envelope theorem, the derivative of C with respect to Q is,

$$\frac{\partial C}{\partial Q} = \frac{1}{A} \times \lambda, \tag{74}$$

where λ is the Lagrange multiplier on the output constraint.

From the envelope theorem, the derivative of C with respect to K is,

$$\frac{\partial C}{\partial K} = -\lambda F_K(Z, K) = -A \times \frac{\partial C}{\partial Q} \times \times F_K(Z, K) \tag{75}$$

Differentiating $\partial C / \partial K$ with respect to Q gives,

$$\frac{\partial C}{\partial K \partial Q} = -A \times \frac{\partial^2 C}{\partial Q^2} \times F_K(Z, K) - A \times \frac{\partial C}{\partial Q} \times F_{ZK}^{\top} \frac{\partial Z}{\partial Q} \tag{76}$$

By Assumption 2:

- Marginal cost is increasing so the first term is negative.
- $F_{ZK} \geq 0$ and inputs are normal so they have increasing conditional demand functions, $\partial Z / \partial Q \geq 0$. So the second term is negative

Therefore,

$$\frac{\partial C}{\partial K \partial Q} \leq 0. \tag{77}$$

□

B.2 Proof of Theorem 1

Proof of Theorem 1.

Assume there is an interior solution $(Q^{\star})$ to the revenue maximization problem. The first order condition then requires,

$$\mathrm{MR}(Q^{\star}) = 0. \tag{78}$$

The second order condition requires that,

$$\mathrm{MR}'(Q^{\star}) \leq 0. \tag{79}$$

Because there is no Q such that $\text{MR}(Q) = MR'(Q) = 0$, it must be that $\text{MR}'(Q^\star) < 0$. Therefore,

$$\text{MR}'\left(Q^{star}\right) \times Q + \text{MR}(Q) < 0, \tag{80}$$

which contradicts Assumption 4. Therefore, there is no interior solution to the revenue maximization problem. □

B.3 Proof of Theorem 2

Proof of Theorem 2.

First, I prove the result when K is capacity. To establish the comparative static result, I make use of the Topkis (1978). The theorem requires that the constraint set is a lattice (closed under minimization and maximization operations), that the objective function is supermodular in all the choice variables, and that it has increasing differences in any given choice with any parameter that we are trying to show the choice is increasing in. What this amounts to is showing that the objective function has positive pairwise cross partials with respect to any two choices and any one choice and any one parameter.

These requirements are cardinal assumptions. They may hold for some increasing transformations of the choices and parameters but not others. That is the case in this problem. I consider optimization not in the natural parameters but in $\widetilde{Q}_t = Q_t/A_t$, K_t, and A_t. Furthermore, I split the problem into an inner and outer optimization problem and apply the Topkis (1978) theorem to both. I do so because the constraint set $\widetilde{Q}_t A_t \leq K_t$ is *not* a lattice in $\left(\widetilde{Q}_t, A_t, K_t\right)$ (consider $(4,1,4)$ and $(1,2,3)$, the maximum of both is $(4,2,4)$ which is not in the set). But the constraint set *is* a lattice in $\left(\widetilde{Q}_t, K_t\right)$.

I write the optimization problem as,

$$\max_{A} \max_{K,\widetilde{Q}} \quad \sum_{t=1}^{T} \beta_t \times \left[\underbrace{P\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t}_{\text{Revenue}} - \underbrace{C\left(\widetilde{Q}_t, W_t\right)}_{\text{Production costs}} - \underbrace{M\left(A_t, A_{t-1}, \lambda_t\right)}_{\text{Technology cost}} - \underbrace{G_t\left(K_{t+1}, K_t\right)}_{\text{Capital adjustment costs}} \right] \tag{81}$$

$$\text{ST:} \quad \widetilde{Q}_t A_t \leq K_t \tag{82}$$

First, I consider the interior optimization problem. I want to show that $\widetilde{Q}_t\left(s, A^T, W^T\right)$ and $K_t\left(s, A^T, W^T\right)$ are increasing in the strong set order in $\left(s, W^T, A^T\right)$. This requires showing the pairwise partial derivatives are positive.

Derivative with respect to $\widetilde{Q}_t$. The derivative with respect to $\widetilde{Q}_t$ is:

$$\beta_t \times \left[P'\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t^2 + P\left(\widetilde{Q}_t A_t, \xi_t\right) A_t - \frac{\partial C}{\partial \widetilde{Q}_t}\left(\widetilde{Q}_t, K_t, W_t\right) \right] \tag{83}$$

I show that it is increasing in K_t and all elements of (s, A_t).

The derivative with respect to A_t is,

$$\beta_t \times \left[P''\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t^2 A_t^2 + 2P'\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t + P'\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t + P\left(\widetilde{Q}_t A_t, \xi_t\right)\right] \quad (84)$$

$$= \beta_t \frac{\partial}{\partial Q_t} \left[\mathrm{MR}\left(Q_t, \xi_t\right) Q_t\right] \geq 0, \quad (85)$$

by Assumption 4.

The derivative with respect to ξ_t is,

$$\beta_t \times A_t \frac{\partial}{\partial \xi_t} \left[P'\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t + P\left(\widetilde{Q}_t A_t, \xi_t\right)\right] = \beta_t A_t \frac{\partial}{\partial \xi_t} \mathrm{MR}\left(Q_t, \xi_t\right) \geq 0, \quad (86)$$

by Assumption 1.

The derivative with respect to K_t is,

$$-\beta_t \frac{\partial C}{\partial K_t \partial \widetilde{Q}_t} \left(\widetilde{Q}_t, K_t, W_t\right) \geq 0, \quad (87)$$

by Assumption 2.

The derivative with respect to $-W_t$ is,

$$\beta_t \times \frac{\partial C}{\partial \widetilde{Q}_t \partial W_t} \geq 0, \quad (88)$$

by Assumption 2. By the envelope theorem, the derivative with respect to W_t is the conditional input demand function and by Assumption 2, that function is increasing in $\widetilde{Q}_t$.

No other parameters or choices appear in the derivative with respect to $\widetilde{Q}_t$. So all cross partials are positive as required.

Derivative with respect to K_t. The derivative of the objective function with respect to K_t is,

$$-\beta_t \times \frac{\partial G_t}{\partial K_t} \left(K_{t+1}, K_t\right) - \beta_{t-1} \frac{\partial G_{t-1}}{\partial K_t} \left(K_t, K_{t-1}\right) \quad (89)$$

The derivative with respect to K_{t+1} is,

$$-\beta_t \frac{\partial G_t}{\partial K_{t+1} \partial K_t} \left(G_{t+1}, G_t\right) \geq 0, \quad (90)$$

by Assumption 3.

The derivative with respect to K_{t-1} is,

$$-\beta_{t-1} \frac{\partial G_{t-1}}{\partial K_{t-1} \partial K_t} \geq 0, \quad (91)$$

by Assumption 3.

No other parameters or choices appear in the K_t derivative so all cross-partials are positive as required.

By the Topkis (1978) theorem, $\widetilde{Q}_t\left(s, A^T, W^T\right)$ and $K_t\left(s, A^T, W^T\right)$ are increasing in $\left(s, A^T, W^T\right)$.

Now, we move to the outer optimization problem. My goal here is to show that $A_t\left(s, -W^T\right)$ is increasing in $\left(s, -W^T\right)$. By the envelope theorem, the derivative with respect to A_t is,

$$\beta_t \times \left[P'\left(\widetilde{Q}_t A_t, \xi_t\right) \times \widetilde{Q}_t^2 A_t + P\left(\widetilde{Q}_t A_t, \xi_t\right) \times \widetilde{Q}_t - \frac{\partial M}{\partial A_t}\left(A_t, A_{t-1}, \lambda_t\right)\right] \tag{92}$$

$$-\beta_{t+1} \times \frac{\partial M}{\partial A_t}\left(A_{t+1}, A_t, \lambda_{t+1}\right) - \eta_t \widetilde{Q}_t \tag{93}$$

From the first order condition with respect to $\widetilde{Q}_t$,

$$\beta_t \times \left[P' \times \widetilde{Q}_t A_t^2 + P \times A_t - C_{\widetilde{Q}}\right] = \eta_t A_t \tag{94}$$

$$\implies \beta_t \times \left[Q_t \mathrm{MR}\left(Q_t, \xi_t\right) - \widetilde{Q}_t \times C_{\widetilde{Q}}\right] = \widetilde{Q}_t A_t \eta_t \tag{95}$$

Substituting this expression in to the derivative with respect to A_t gives,

$$\beta_t \left[\frac{\widetilde{Q}_t}{A_t} \times \frac{\partial C}{\partial \widetilde{Q}_t}\left(\widetilde{Q}_t, W_t\right) - \frac{\partial M}{\partial A_t}\left(A_t, A_{t-1}, \lambda_t\right)\right] \tag{96}$$

$$-\beta_{t+1} \times \frac{\partial M}{\partial A_t}\left(A_{t+1}, A_t, \lambda_{t+1}\right) \tag{97}$$

The derivative with respect to A_{t+1} gives,

$$\beta_t \left[\frac{\frac{\partial \widetilde{Q}_t}{\partial A_{t+1}}}{A_t} \times \frac{\partial C}{\partial \widetilde{Q}_t}\left(\widetilde{Q}_t, W_t\right) + \frac{\widetilde{Q}_t}{A_t} \times \frac{\partial^2 C}{\partial \widetilde{Q}_t^2}\left(\widetilde{Q}_t, W_t\right) \times \frac{\partial \widetilde{Q}_t}{\partial A_{t+1}}\right] \tag{98}$$

$$-\beta_{t+1} \frac{\partial M}{\partial A_{t+1} \partial A_t}\left(A_{t+1}, A_t, \lambda_t\right) \tag{99}$$

The first term is positive because $\widetilde{Q}_t$ is increasing in A_{t+1} and because marginal cost is positive. By Assumption 2, marginal cost is increasing in output, so the second term is positive as well. The third term is positive by Assumption 3.

The derivative with respect to A_{t-1} is, symmetrically, also positive.

For ξ_t, it is easier algebraically to first use the envelope theorem to recover the derivative of the objective with respect to ξ_t and then take the derivative with respect to $A_{t'}$ (where $t' \in \{1, \ldots, T\}$). The derivative with respect to ξ_t is,

$$\beta_t \frac{\partial P}{\partial \xi} \times \widetilde{Q}_t A_t. \tag{100}$$

Differentiating with respect to $A_{t'}$ gives,

$$\beta_t \times \left[\frac{\partial P}{\partial Q \partial \xi} \times \frac{\partial Q_t}{\partial A_{t'}} \times Q_t + \frac{\partial P}{\partial \xi} \times \frac{\partial Q_t}{\partial A_{t'}}\right] = \tag{101}$$

$$\beta_t \times \frac{\partial Q_t}{\partial A_{t'}} \times \left[\frac{\partial P}{\partial Q \partial \xi} \times Q_t + \frac{\partial P}{\partial \xi}\right] = \tag{102}$$

$$\beta_t \times \frac{\partial Q_t}{\partial A_{t'}} \times \frac{\partial}{\partial \xi} \mathrm{MR}\left(Q, \xi\right) \geq 0, \tag{103}$$

by Assumption 1 and by the previous result that $\widetilde{Q}_t$ is increasing in $A_{t'}$.

The derivative of the objective with respect to $-W_t$ gives,

$$\beta_t \frac{\partial C}{\partial W_t}\left(\widetilde{Q}_t, W_t\right) \tag{104}$$

Differentiating with respect to $A_{t'}$ gives,

$$\beta_t \frac{\partial C}{\partial \widetilde{Q}_t \partial W_t}\left(\widetilde{Q}_t, W_t\right) \times \frac{\partial \widetilde{Q}_t}{\partial A_{t'}} \geq 0, \tag{105}$$

by Assumption 2 and the result that $\widetilde{Q}_t$ is increasing in $A_{t'}$.

Therefore, $A_t\left(s, -W^T\right)$, $Q_t\left(s, -W^T\right)$, and $K_t\left(s, -W^T\right)$ are increasing in all arguments.

K is capital. When K is capital, not capacity, then the plant's problem is,

$$\max_{K, A, \widetilde{Q}} \sum_{t=1}^{T} \beta_t \times \left[P\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t A_t - C\left(\widetilde{Q}_t, K_t, W_t\right) - M\left(A_t, A_{t-1}, \lambda_t\right) - G_t\left(K_{t+1}, K_t\right)\right], \tag{106}$$

and the constraint set is simply $\left(\widetilde{Q}_t, A_t, K_t\right) \geq 0$ which is a lattice. I show that $\widetilde{Q}_t\left(s, -W^T\right)$, $K_t\left(s, -W^T\right)$, and $A_t\left(s, -W^T\right)$ are increasing in s for fixed W^T. The proof strategy is similar to the above.

Derivative with respect to $\widetilde{Q}_t$. The derivative with respect to $\widetilde{Q}_t$ is the same as in the previous proof. It is increasing in (A_t, K_t) and in s as required.

Derivative with respect to K_t. The derivative with respect to K_t is the same as in the previous proof. It is increasing in $\left(\widetilde{Q}_t, A_t, K_{t-1}, K_{t+1}\right)$ and in s.

Derivative with respect to A_t. The derivative with respect to A_t is,

$$\beta_t \times \left[P'\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t^2 A_t + P\left(\widetilde{Q}_t A_t, \xi_t\right) \widetilde{Q}_t - \frac{\partial M}{\partial A_t}\left(A_t, A_{t-1}, \lambda_t\right)\right] - \beta_{t+1} \frac{\partial M}{\partial A_t}\left(A_{t+1}, A_t, \lambda_{t+1}\right). \tag{107}$$

It is increasing in $\widetilde{Q}_t$ because we have already shown that $\widetilde{Q}_t$'s derivative is increasing in A_t and increasing in K_t because we have already shown that K_t's derivative is increasing in A_t.

The derivative with respect to ξ_t is,

$$\beta_t \times \widetilde{Q}_t \times \frac{\partial}{\partial \xi_t}\left[P'\left(\widetilde{Q}_t A_t, \xi_t\right)\widetilde{Q}_t A_t + P\left(\widetilde{Q}_t A_t, \xi_t\right)\right] \geq 0, \tag{108}$$

by Assumption 1.

The derivative with respect to A_{t-1} is,

$$-\beta_t \frac{\partial M}{\partial A_{t-1}\partial A_{t-1}}\left(A_t, A_{t-1}, \lambda_t\right) \geq 0, \tag{109}$$

by Assumption 3.

The derivative with respect to λ_t is,

$$-\beta_t \frac{\partial M}{\partial \lambda_t \partial A_t} \geq 0, \tag{110}$$

by Assumption 1.

The derivative with respect to A_{t+1} is,

$$-\beta_{t+1} \frac{\partial M}{\partial A_{t+1}\partial A_t}\left(A_{t+1}, A_t, \lambda_{t+1}\right) \geq 0 \tag{111}$$

The derivative with respect to λ_{t+1} is,

$$-\beta_{t+1} \frac{\partial M}{\partial \lambda_{t+1}\partial A_t} M\left(A_{t+1}, A_t, \lambda_{t+1}\right) \geq 0, \tag{112}$$

by Assumption 1.

Which establishes that the objective function is supermodular in $\left(\widetilde{Q}, A, K\right)$ and has increasing differences between s and these choices. Therefore, by Topkis (1978), $A_t\left(s, -W^T\right)$, $Q_t\left(s, -W^T\right)$, and $K_t\left(s, -W^T\right)$ are increasing in s.

□

B.4 Proof of Lemma 2

Proof of Lemma 2.

σ_t is an increasing function of $(\widetilde{s}_{t-1}, \upsilon_t)$. υ_t and $\widetilde{s}_{t-1}$ are both positively associated and υ_t is independent of $\widetilde{s}_{t-1}$. By Property 1, we know that the random vector $(\widetilde{s}_{t-1}, \upsilon_t)$ is positively associated. By Property 2, we know that because $\widetilde{s}_t = \sigma_t\left(\widetilde{s}_{t-1}, \upsilon_t\right)$ is a vector of increasing function of that positively associated random vector, it too is positively associated. So $\widetilde{s}_t$ is positively associated. □

B.5 Proof of Theorem 3

Proof of Theorem 3.

By Lemma 2, if $\widetilde{s}_{t-1}$ is positively, then so is $\widetilde{s}_t$.

By Assumption 7, $\widetilde{s}_0$ is positively associated.

By induction, $\widetilde{s}_t$ is positively associated for all t. $\widetilde{s} = \cup_{t=0}^{T} \widetilde{s}_t$. Because $(\widetilde{s}_0, \widetilde{s}_1)$ is positively associated, it follows that $(\widetilde{s}_2, \widetilde{s}_1, \widetilde{s}_0)$ is positively associated because $\widetilde{s}_2$ is an increasing function of $(\widetilde{s}_1, \widetilde{s}_0, \upsilon_2)$ which is positively associated. Following the induction argument, $\widetilde{s}$ itself is positively associated. □

B.6 Proof of Theorem 5

Proof of Theorem 5.

The derivative of f^α is,

$$Df^\alpha = (1-\alpha)\, Df^0 + \alpha Df^1 \geq 0, \tag{113}$$

because $Df^0 \geq 0$ and $Df^1 \geq 0$.

For any two increasing functions $\phi_1(q,k)$ and $\phi_2(q,k)$,

$$\operatorname{cov}\left[\left(q - f^0\right)\phi_1, \phi_2\right] \geq 0, \quad \operatorname{cov}\left[\left(q - f^1\right)\phi_1, \phi_2\right] \geq 0 \tag{114}$$

$$\implies \alpha\operatorname{cov}\left[\left(q - f^0\right)\phi_1, \phi_2\right] + (1-\alpha)\operatorname{cov}\left[\left(q - f^1\right)\phi_1, \phi_2\right] \geq 0 \tag{115}$$

$$\implies \operatorname{cov}\left[\left(q - f^\alpha\right)\phi_1, \phi_2\right] \geq 0 \tag{116}$$

So $f^\alpha \in \mathcal{F}_{\text{LPA}}$ because it satisfies both linear positive association and is an increasing function. □

B.7 Proof of Theorem 6

Proof of Theorem 6.

Call the set in the theorem's conclusion which only uses a finite number of ϕ functions $\widetilde{\Theta}_{\text{CD,LPA}}$. To show $\widetilde{\Theta}_{\text{CD,LPA}} = \Theta_{\text{CD,LPA}}$, I show that $\widetilde{\Theta}_{\text{CD,LPA}}$ both contains and is contained by $\Theta_{\text{CD,LPA}}$.

Because $\left(q - \underline{q}\right)$ and $(k - \underline{k})$ are both functions that would be used in $\Theta_{\text{CD,LPA}}$ by making the choices $\varphi_1 = -\underline{q}$, $\varphi_q = 1$, $\varphi_k = 0$ and $\varphi_1 = -\underline{k}$, $\varphi_q = 0$, $\varphi_k = 1$, clearly any element of θ that belongs to $\Theta_{\text{CD,LPA}}$ will also belong to $\widetilde{\Theta}_{\text{CD,LPA}}$.

I show that any element $\theta \in \widetilde{\Theta}_{\text{CD,LPA}}$ also belongs to $\Theta_{\text{CD,LPA}}$. Let $\theta \in \widetilde{\Theta}_{\text{CD,LPA}}$, then we know:

$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(q-\underline{q}\right),\left(q-\underline{q}\right)\right)\geq 0 \tag{117}$$
$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(q-\underline{q}\right),\left(k-\underline{k}\right)\right)\geq 0 \tag{118}$$
$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(k-\underline{k}\right),\left(k-\underline{k}\right)\right)\geq 0 \tag{119}$$
$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(k-\underline{k}\right),\left(q-\underline{q}\right)\right)\geq 0 \tag{120}$$

We want to show that θ satisfies,

$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left[\varphi_{1,1}+\varphi_{1,k}k+\varphi_{1,q}q\right],\varphi_{2,1}+\varphi_{2,k}k+\varphi_{2,q}q\right)\geq 0 \tag{121}$$

for φ that satisfy: $\varphi_k\geq 0$, $\varphi_q\geq 0$ and $\varphi_1+\varphi_k k+\varphi_q q\geq 0$.

First, consider only φ such that $\varphi_1=-\varphi_k\underline{k}-\varphi_q\underline{q}$. Then the functions are: $\varphi_q\left(q-\underline{q}\right)+\varphi_k\left(k-\underline{k}\right)$. I show that θ satisfies the covariance restrictions for functions of this form:

- θ satisfies the covariance restrictions for functions of the form $\varphi_{1,q}\left(q-\underline{q}\right)$ and $\varphi_{2,q}\left(q-\underline{q}\right)$,

$$\varphi_{1,q}\varphi_{2,q}\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(q-\underline{q}\right),\left(q-\underline{q}\right)\right) \tag{122}$$
$$=\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\varphi_{1,q}\left(q-\underline{q}\right),\varphi_{2,q}\left(q-\underline{q}\right)\right)\geq 0 \tag{123}$$

- θ satisfies the covariance restrictions for functions of the form $\varphi_{1,k}\left(k-\underline{k}\right)$ and $\varphi_{2,q}\left(q-\underline{q}\right)$,

$$\varphi_{1,k}\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(k-\underline{k}\right),\varphi_{2,q}\left(q-\underline{q}\right)\right) \tag{124}$$
$$=\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\varphi_{1,k}\left(k-\underline{k}\right),\varphi_{2,q}\left(q-\underline{q}\right)\right)\geq 0 \tag{125}$$

- Summing the two restrictions shows that θ satisfies the covariance restrictions for functions of the form $\varphi_{1,q}\left(q-\underline{q}\right)+\varphi_{1,k}\left(k-\underline{k}\right)$ and $\varphi_{2,q}\left(q-\underline{q}\right)$,

$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\varphi_{1,k}\left(k-\underline{k}\right),\varphi_{2,q}\left(q-\underline{q}\right)\right) \tag{126}$$
$$+\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\varphi_{1,q}\left(q-\underline{q}\right),\varphi_{2,q}\left(q-\underline{q}\right)\right) \tag{127}$$
$$=\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\varphi_{2,q}\left(q-\underline{q}\right)\right)\geq 0 \tag{128}$$

- Symmetrically, we can show that θ satisfies the covariance restrictions for functions of the form $\varphi_{1,q}\left(q-\underline{q}\right)+\varphi_{1,k}\left(k-\underline{k}\right)$ and $\varphi_{2,k}\left(k-\underline{k}\right)$.
- Summing the covariance restrictions together establishes that θ satisfies all covariance restrictions of the form $\varphi_{1,q}\left(q-\underline{q}\right)+\varphi_{1,k}\left(k-\underline{k}\right)$ and $\varphi_{2,q}\left(q-\underline{q}\right)+\varphi_{2,k}\left(k-\underline{k}\right)$,

$$\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\varphi_{2,q}\left(q-\underline{q}\right)\right) \tag{129}$$
$$+\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\varphi_{2,k}\left(k-\underline{k}\right)\right) \tag{130}$$
$$=\text{cov}\left(\left[q-\theta_1-z^\top\theta_z\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\right. \tag{131}$$
$$\left.\varphi_{2,q}\left(q-\underline{q}\right)+\varphi_{2,k}\left(k-\underline{k}\right)\right)\geq 0 \tag{132}$$

For any function $\varphi_1 + \varphi_q q + \varphi_k k$ such that $\varphi_q \geq 0$ and $\varphi_k \geq 0$, the minimum value of the function across observed (q, k) is no smaller than $\varphi_1 + \varphi_q \underline{q} + \varphi_k \underline{k}$. Because the function must be positive, $\varphi_1 \geq -\varphi_q \underline{q} - \varphi_k$. We have already shown that θ satisfies the covariance restrictions for $\varphi_1 = -\varphi_q \underline{q} - \varphi_k \underline{k}$.

Clearly, the constant $\varphi_{2,1}$ does not matter for the covariance because $\operatorname{cov}(X_1, X_2 + \text{constant}) = \operatorname{cov}(X_1, X_2)$. So the covariance restrictions hold for $\varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k$.

Because $\varphi_{1,1} = -\varphi_{1,q}\underline{q} - \varphi_{1,k}\underline{k} + \Delta$ where $\Delta \geq 0$, by the linearity of covariance, we can write the covariance restrictions for $\varphi_{1,1} + \varphi_{1,q} q + \varphi_{1,k} k$ as,

$$\operatorname{cov}\left(\left[q - \theta_1 - z^\top \theta_z\right] \Delta,\right. \tag{133}$$

$$\left.\varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k\right) \tag{134}$$

$$+\operatorname{cov}\left(\left[q - \theta_1 - z^\top \theta_z\right] \left(\varphi_{1,k}\left(k - \underline{k}\right) + \varphi_{1,q}\left(q - \underline{q}\right)\right),\right. \tag{135}$$

$$\left.\varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k\right). \tag{136}$$

The first term is positive because $\theta \in \widetilde{\Theta}_{\text{CD,LPA}}$ and $\Delta \geq 0$ so $\Delta \times \operatorname{cov}(a, \varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k) \geq 0$.

The second term is positive by the above results establishing that the linear positive association holds for φ of that form for $\theta \in \widetilde{\Theta}_{\text{CD,LPA}}$. So,

$$\operatorname{cov}\left(\left[q - \theta_1 - z^\top \theta_z\right] (\varphi_{1,1} + \varphi_{1,k} k + \varphi_{1,q} q),\right. \tag{137}$$

$$\left.\varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k\right) \geq 0. \tag{138}$$

Therefore, $\theta \in \Theta_{\text{CD,LPA}}$ because θ satisfies the covariance restrictions for any increasing, positive function of the form: $\varphi_1 + \varphi_q q + \varphi_k k$.

This establishes that $\widetilde{\Theta}_{\text{CD,LPA}} = \Theta_{\text{CD,LPA}}$. □

B.8 Proof of Theorem 7

Proof of Theorem 7.

The parameter set defined by the intersection only over the step functions contains the parameter set identified by the linear positive association because the step functions are increasing functions and the linear positive association is the intersection over all increasing functions. I show it is also contained by the set identified by linear positive association, establishing equality of the two sets.

The proof is related to Theorem 3.4 of Esary, Proschan, and Walkup (1967).

Let $t_1, \ldots, t_k$ be a finite set of scalars which approach the rational numbers in $[0, 1]$ as $k \to \infty$.

Let ϕ_1 and ϕ_2 be two non-negative, bounded, continuous, increasing functions that belong to L_1.

Define $h_k(z) = \phi_1(\eta)$ where,

$$\eta_i = \varphi_i^{-1}\left(\max_{j\in\{1,\dots,k\}}\{t_j : \mathbf{1}(\varphi_i(z_i) \geq t_j) = 1\}\right) \text{ for } i = 1, \dots, L. \tag{139}$$

Define $v_k(z) = \phi_2(\eta)$.

Assume,

$$\operatorname{cov}(a\mathbf{1}(\varphi(z_i) \geq u), \mathbf{1}(\varphi(z_i) \geq v)) \geq 0. \tag{140}$$

For all u and v in the rational numbers between $[0, 1]$.

There is a positive function κ_h such that,

$$h_k(z) = \sum_{t_1=1}^{k} \cdots \sum_{t_L=1}^{k} \kappa_h(t_1, \dots, t_L) \times \mathbf{1}(\varphi_1(z_1) \geq t_1, \dots, \varphi_L(z_L) \geq t_L). \tag{141}$$

There is such a κ_v for v_k as well.

Therefore, by the linearity of the covariance,

$$\sum_{t^1}\sum_{t^2} \kappa_h\left(t^1\right)\kappa_v\left(t^2\right)\operatorname{cov}\left(a\mathbf{1}\left(\varphi(z) \geq t^1\right), \mathbf{1}\left(\varphi(z) \geq t^2\right)\right) = \tag{142}$$

$$\operatorname{cov}(ah_k(z), v_k(z)) \geq 0. \tag{143}$$

Because ϕ_1 and ϕ_2 are continuous, $0 \leq h^k \to \phi_1$, $0 \leq v^k \to \phi_2$, and $0 \leq v^k h^k \to \phi_1\phi_2$ point-wise (as $k \to \infty$, as $\{t_1, \dots, t_k\} \to \mathbb{Q} \cap [0, 1]$). Because ϕ_1 and ϕ_2 are bounded and because the sequences in k are increasing, by the monotone convergence theorem,

$$\mathbb{E}(ah_k(z)) \to \mathbb{E}(a\phi_1(z)) \tag{144}$$

$$\mathbb{E}(v_k(z)) \to \mathbb{E}(\phi_2(z)) \tag{145}$$

$$\mathbb{E}(ah_k(z)v_k(z)) \to \mathbb{E}(a\phi_1(z)\phi_2(z)). \tag{146}$$

So I have what I wanted to show,

$$0 \leq \lim_{k\to\infty} \operatorname{cov}(ah_k(z), v_k(z)) = \operatorname{cov}(a\phi_1(z), \phi_2(z)). \tag{147}$$

□