Identifying productivity when it is a choice

Zach Flynn*

February 26, 2019

Abstract

Productivity is typically modeled as an exogenous plant characteristic, but the people who run the plant make choices that affect how productive it is. Instruments that are uncorrelated with productivity but correlated with input use are therefore rare because when plants want to adjust one factor of production, input use, they also want to adjust other factors of production, like productivity. This article develops a method to partially identify statistics of the productivity distribution based on a monotone comparative static result without assuming productivity is exogenous or requiring an instrumental variable. I use the method to study the productivity effect of restructuring 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

Productivity is why plants that use the same inputs produce different levels of output. It is typically modeled as an exogenous characteristics of the plant[1], but plants also 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 all make choices daily in response to the incentives they face that affect the productivity of the plant. This article develops a method to partially identify productivity when it is determined by these and other unobserved factors of production.

To be concrete, suppose productivity is input neutral,

$$\log \text{Output} = f(\text{Inputs}) + a, \tag{1}$$

where a is log total factor productivity and $f(\cdot)$ is the log production function. We cannot identify productivity simply by regressing log output on the inputs because productivity is correlated with input use. Greater productivity increases the marginal product of input use, changing the plant's optimal input choice. This is the basic problem of identifying the production function, and it must be dealt with whether productivity is a choice or not.

When productivity is endogenous, instruments to resolve this identification problem are more difficult to find than when productivity is exogenous. Instrumental variable techniques require an observed variable that is correlated with input use but not productivity. But because both inputs and productivity determine output, variables that affect the plant's output choice will affect both its choice of inputs *and* productivity[2]. For example, demand shocks, a common instrument in exogenous productivity models, affect output choice, and so, the choice of *both* inputs and productivity when it is endogenous. Plants want to adjust their productivity when they want to adjust their input use.

I argue changes in the plant's environment affect productivity choice and plant size in the *same way*. I build a general model of production and prove a comparative static result: when a change in the plant's state variables increases the plant's optimal output choice, the plant also chooses greater productivity. This result and a statistical assumption on the plant's state variables together imply a restriction on the production function that can be used to partially identify statistics of the productivity distribution without making strong assumptions about how the plant chooses its productivity.

[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. 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 called it, "the measure of our ignorance". The substantial identification problems this view of productivity introduces are the subject of this article.

[2] Without making information and timing restrictions like an assumption that the shocks that affect output choice are not known to the plant until after it has chosen its productivity.

The production function identification problem has been discussed in a large and old literature[3] 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 (2015). The proxy method generates instruments for input choice by making information and timing assumptions about when plants make decisions. These assumptions allow lagged input choices to proxy for past productivity[4]. The method assumes productivity is exogenous or that choices that affect productivity can be controlled for by observables (see Doraszelski and Jaumandreu 2013). It 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.

Exogenous productivity models restrict the class of questions we can answer. In particular, estimates recovered from these models cannot be used to learn whether a policy *incentivized* plants to become more productive. Incentives can only affect productivity if plants can adjust their productivity. More critically, it is simply likely the case that productivity is not, in fact, exogenous so exogenous productivity models may be misspecified.

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). They proposed a modification of the proxy model that allows productivity to be endogenous when it is controlled by *observed* choices[5]. But in most datasets, variables that plausibly control for productivity choice are difficult to find. The fact that productivity itself is unobserved suggests investment in it will be difficult to observe as well. These methods also assume any observed investment in productivity is made before unobserved, exogenous productivity shocks are known to the plant.

An alternative method to the one proposed in this paper would be to develop instruments by making timing assumptions about when plants choose their productivity. We could assume the choice to invest in productivity is made before some observed shock that affects input use. What this would do is "divide up" the state space so that there are shocks that affect input choice but not productivity. We could then use the shocks to instrument for input choice. But, to be clear, the plant will *want* to adjust its productivity. This approach would simply assume they are *unable* to do so, a difficult assumption to justify without precise knowledge of what productivity, an unobservable, is.

[3]This article is also related to the macroeconomics literature where firms endogenously choose production techniques, see Jones (2005) and Uras and Wang (2017), and the literature estimating structural models of principal agent problems, see Wolak (1994).

[4]The underlying economic model behind the proxy approach is similar to the models of firm and industry dynamics developed in Jovanovic (1982) and Hopenhayn (1992).

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

I allow plants to choose their productivity without requiring that we observe the investment in productivity and without making such a timing assumption. I develop a partial identification strategy based on a robust comparative static result I find in Section 2.

In Section 3, I show how to make inference on coefficients in regressions where log productivity is the dependent variable. Regressions of productivity on policy variables and plant characteristics are a standard statistic of interest in the productivity literature. 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 simply by drawing from the posterior of a set of reduced-form parameters and re-solving the linear programs repeatedly, a trivial computation compared to the usual grid search required for inference on partially identified parameters[6].

In Section 4, 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-restructuring may also contribute to lower productivity choice post-restructuring. I discuss these effects in more detail in Section 4.1.

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 4.3. I find restructuring caused power plants to lower their fuel productivity by between 1.16% 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.

[6]See Chernozhukov, Lee, and Rosen (2013), Chernozhukov, Hong, and Tamer (2007), and many similar examples.

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, as discussed above, demand is only uncorrelated with productivity if productivity is exogenous. 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 4.3.1.

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

My goal is to 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. To this end, my identification strategy is not based on first order conditions but on a robust comparative static result. I say the comparative static is robust because it holds in a wide variety of specializations of a general model I develop. It holds when productivity is a static decision[7] and when it is chosen with dynamic considerations[8], 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.

Intuitively, I find productivity and output are chosen in the same *way*. When plants want to choose greater output, they want to be more productive (Section 2.1). This result paired with a statistical assumption on the unobserved state variables (Section 2.2) leads to the conclusion that productivity is *positively associated* with output and capacity in the sense of Esary, Proschan, and Walkup (1967). I use this result to partially identify the production function and the productivity distribution (Section 2.3).

2.1 The model and a comparative static result

Let lowercase variables be in logs when uppercase variables are levels (for example, $\log Q = q$). The production equation is,

$$Q = F(Z, K)\, A \implies q = f(z, k) + a, \tag{2}$$

[7]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.

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

where Q is output, Z is a vector of L variable inputs, K is a capital or capacity input, and A is total factor productivity.

In the model, plants choose their output Q, their capacity K, and their productivity A to solve the following problem for a sequence of discount factors $\beta_t \geq 0$,

$$\max_{K,A,Q} \sum_{t=1}^{T} \beta_t \times \left[\underbrace{P\left(Q_t, \xi_t\right) Q_t}_{\text{Revenue}} - \underbrace{C\left(\frac{Q_t}{A_t}, K_t, W_t\right)}_{\text{Production costs}} - \underbrace{M\left(A_t, A_{t-1}, \lambda_t\right)}_{\text{Technology cost}} - \underbrace{G\left(K_{t+1}, K_t\right)}_{\text{Capital adjustment costs}} \right] \tag{3}$$

Subject to the capacity and nonnegativity constraints:

$$0 \leq Q_t \leq K_t, \quad A_t \geq 0. \tag{4}$$

I describe each element of the model in turn:

- $P(Q, \xi)$ is an industry-level function which gives the (residual) demand function for each plant. ξ varies across plants and indexes how the (residual) demand function varies for each plant.
- C is the variable cost function. Let Z be the variable inputs, K be an input that determines capacity, W be input prices (which may vary by plant), and A be total factor productivity,

$$C\left(\frac{Q}{A}, K, W\right) = \min W^{\top} Z \quad \text{st:} \quad F(Z, K) \geq \frac{Q}{A} \text{ if } Q \leq K \tag{5}$$

$$= \infty \text{ if } Q > K. \tag{6}$$

  Capacity constraints are not important for my results, but capacity plays a role in my empirical analysis so I introduce it now.

- M is the function that 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 gives the cost of building a certain level of capacity given current capacity, so that it is more expensive to go from low to high capacity than from medium to high capacity.
- For simplicity, I assume plants believe that they have perfect foresight. Plants look T periods into the future when making their decisions and solve the optimization problem in (3) subject to the capacity constraints (4).

The model nests the basic proxy model used in current structural production function estimators where productivity is exogenous[9], competition is perfect, and there are no capacity constraints. But the model is more general. It has productivity choice, capacity constraints, and plant-level variation in the residual demand curve.

I make the following assumptions to give meaning to the ξ and λ parameters and provide a way to order the demand and cost of productivity functions:

Assumption 1. Marginal revenue is increasing in ξ,

$$\frac{\partial}{\partial \xi}\left[\frac{\partial P}{\partial Q}(Q,\xi)\,Q + P(Q,\xi)\right] \geq 0. \tag{7}$$

Assumption 2. The marginal cost of productivity is decreasing in λ,

$$\frac{\partial M}{\partial A_t \partial \lambda}(A_t, A_{t-1}, \lambda) \leq 0, \tag{8}$$

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

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

The main restriction these assumptions put on the model is that the demand and cost of productivity functions can be ordered across plants. There are not plants that have have higher marginal revenues than other plants at low levels of output and lower marginal revenue than other plants at high levels of output. There are not plants that are better than other plants at capitalizing on prior productivity investments but with higher marginal costs of current productivity than other plants.

In the proxy model of Olley and Pakes (1996) and Levinsohn and Petrin (2003), the plants can be ordered across a single dimension, productivity. So this model is more flexible in the sense that the plants can only be partially ordered. It is possible to have a good demand curve and bad costs of productivity, but I still require some ordering of the plant's unobserved characteristics for the comparative static result to hold.

I also make assumptions on the functions that define the plant's problem: $P(\cdot,\cdot)$, $C(\cdot,\cdot,\cdot)$, $M(\cdot,\cdot)$, and $G(\cdot,\cdot)$.

Assumptions on C, the variable cost function. I assume that if capacity affects variable costs

[9] 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 2). Then, λ_t is the exogenous productivity sequence. Exogenous productivity models are endogenous productivity models where it is infinitely costly to adjust productivity.

(for $Q < K$), then it must reduce marginal cost. Intuitively, greater K reduces the amount of additional variable inputs (Z) that would be required to produce additional output, reducing marginal cost.

Assumption 3. Marginal cost of the variable inputs is decreasing in capacity (for $Q < K$),

$$\frac{\partial}{\partial K \partial Q} C\left(\frac{Q}{A}, K, W\right) \leq 0. \tag{10}$$

I call K "capacity" because I allow it to determine capacity, but it could be any form of capital because the capacity constraint is not necessary for the results to hold.

I also assume that the elasticity of the marginal cost of the variable inputs is not too negative. Define $\widetilde{Q} = Q/A$.

Assumption 4. The elasticity of the marginal cost curve with respect to $\widetilde{Q}$ must be greater than or equal to -1,

$$\frac{\partial \log \frac{\partial C}{\partial \widetilde{Q}}\left(\widetilde{Q}, K, W\right)}{\partial \log \widetilde{Q}} \geq -1. \tag{11}$$

While this assumption is about the shape of the production function, it is weak. It requires that when productivity-normalized output increases by one percent, marginal cost falls by less one percent. This restriction is usually required for an interior solution to the optimization problem in any case unless marginal revenue is decreasing at a similarly high rate. The assumption is always true when C has a constant elasticity in $\widetilde{Q}$ and when the production function $F(Z, K)$ is concave in Z for fixed K.

- **Constant elasticity given (K, W) cost functions satisfy Assumption 4**. Temporarily suppress the dependence of C on K and W. Suppose that, for fixed (K, W), $C\left(\widetilde{Q}\right) \propto \widetilde{Q}^{\gamma}$ where $\gamma > 0$ because costs increase in output. The elasticity of the marginal cost curve is then $(\gamma - 1)$. Because $\gamma > 0$, the elasticity of the marginal cost curve must be greater than -1 so Assumption 4 is satisfied, regardless of how increasing or decreasing the returns to scale are. As a corollary of this result, if $F(Z, K) = F_0(K) \prod_{\ell=1}^{L} Z_{\ell}^{\theta_{\ell}(K)}$ (a Cobb Douglas in Z production function), then Assumption 4 is satisfied because the cost function has a constant elasticity in $\widetilde{Q}$.

- **Concave in Z (given K) production functions satisfy Assumption 4.** Define technology (given K) as,

$$\text{TECH} = \left\{\left(\widetilde{Q}, Z\right) : F(Z, K) \geq \widetilde{Q}\right\}. \tag{12}$$

If TECH is a convex set ("convex technology", given K), then the cost function for Z (for fixed K) is convex, so its marginal cost curve will be increasing and satisfy the requirement that the elasticity of the marginal cost curve is greater than or equal to -1. If $F(Z, K)$ is concave in Z for any fixed K, then TECH is convex and the assumption is satisfied. For example, $F = Z^{0.8} K^{0.3}$ would satisfy the assumption.

Assumptions on G, the cost of increasing capacity (capital). I assume the marginal cost of increasing capacity given a certain level of prior capacity in period $t+1$ is (weakly) less if the plant has higher capacity in period t. This allows the input to be dynamic and rules out only the case where having higher capacity previously makes it more difficult to maintain that capacity in the future.

Assumption 5. The cross partial of $G(K_{t+1}, K_t)$ is negative,

$$\frac{\partial G}{\partial K_t \partial K_{t+1}} (K_{t+1}, K_t) \leq 0 \tag{13}$$

Assumption 5 allows, for example, $G(K_{t+1}, K_t) = (K_{t+1} - K_t)^{\text{exponent}}$ with an exponent greater than 1 (convex adjustment costs).

Assumptions on M, the cost of increasing productivity. I assume that the marginal cost of present productivity is not increasing in past productivity. It is less expensive to choose high productivity levels if the plant was highly productive in the past,

Assumption 6. The cross partial of $M(A_t, A_{t-1}, \lambda_t)$ is negative for each λ,

$$\frac{\partial M}{\partial A_t \partial A_{t-1}} \leq 0. \tag{14}$$

Assumption 6, similarly to Assumption 5, allows, for example, $[A_t - A_{t-1}]^{\text{exponent}}$ with an exponent greater than 1 (convex adjustment costs). This assumption allows A_t to be a "capital"-like input, a stock that is accumulated by the plant over time.

Given the above model and its assumptions, I prove a comparative static result (Theorem 1) that forms the basis of the partial identification approach I take to the data.

Theorem 1. Define $W^T = (W_1, \ldots, W_T)$ and $s = \left(A_0, K_0, \{\lambda_t\}_{t=1}^{T}, \{\xi_t\}_{t=1}^{T}\right)$ and let $a_t = a\left(s, -W^T, t\right), q_t = q\left(s, -W^T, t\right), k_t = k\left(s, -W^T, t\right)$ be the optimal productivity, output, and capacity choices given $\left(s, W^T\right)$. Given Assumptions 1, 2, 3, 4, 5, and 6, $a(\cdot, -W, t)$, $q(\cdot, -W, t)$, and $k(\cdot, -W, t)$ are weakly increasing functions for all $(-W, t)$. In other words, productivity, output, and capacity choice are all increasing in s for fixed variable input prices.

Proof. See Appendix B. □

The result holds because the objective function has increasing differences across the choices (Q_t, K_t, A_t) and the parameters s and is supermodular among the choices. See Topkis (1978).

2.2 Markovian state variables and positive association

To turn the comparative static result into an assumption that we can use to make inference on the productivity distribution, I make an intuitive statistical assumption about the distribution of the state variables s. I show a standard Markovian structure on the state variables implies that s is a *positively associated* random vector in the sense introduced by Esary, Proschan, and Walkup (1967).

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

$$\text{cov}\left[\phi_1(X), \phi_2(X)\right] \geq 0, \tag{15}$$

then X is a positively associated random vector.

Esary, Proschan, and Walkup (1967) show positive association has two useful properties that I exploit here:

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.

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

Positive association of the state vector s arises naturally if we assume a standard Markov structure on the state variables. I use the Markov structure in Assumptions 7, 8, and 9.

Assumption 7. Define $s_t = (K_0, A_0, \xi_t, \lambda_t)$. State variables evolve according to the rule,

$$s_{j,t} = \sigma_{j,t}\left(s_{t-1}, \upsilon_{j,t}\right), \tag{16}$$

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

Assumption 8. $(\xi_0, \lambda_0, A_0, K_0)$ is positively associated.

Assumption 9. $\upsilon_t = (\upsilon_{1,t}, \ldots, \upsilon_{J,t})$ is a positively associated vector. Define $s^\tau = \cup_{t=0}^{\tau} s_t$. s^{t-1} and υ_t are independent.

Assumption 7 nests the standard Markov structure, $s_{j,t} = \sigma_{j,t}(s_{j,t-1}, \upsilon_{j,t})$, and allows for larger marginal revenue curves to lead to weakly lower marginal cost of productivity curves.

Assumption 8 says that s_0 is positively associated. It is sufficient that, at some point in the past, $(\xi_0, \lambda_0, A_0, K_0)$ were independent of each other. This initial condition can also be supported by the economics of the model. If:

- (ξ_0, K_0) is positively associated because revenue curves which encourage greater output in the past are likely positively related to having built greater capacity in the past;
- (λ_0, A_0) is positively associated because cost of productivity curves that encourage greater productivity choice are likely correlated with higher productivity;
- and the two sets of parameters (ξ_0, K_0) and (λ_0, A_0) are initially independent,

then Assumption 8 is satisfied.

Assumption 9 adds the assumption that the shocks to the state variables (υ) are either independent of each other[10] or they are positively related. It also adds the Markov assumption that current shocks are independent of prior states.

I can then prove Theorem 2 which concludes that s is positively associated.

Theorem 2. Given Assumption 7, 8, and 9, the state vector s is positively associated.

Proof. See Appendix B. □

Theorem 2 establishes that once s_t is positively associated, s_{t+1} will be positively associated as well. It combines that result with the initial condition to conclude that s itself is positively associated.

Property 2 is the key property of positive association. Increasing functions of positively associated random vectors are themselves positively associated. This property means positive association is an *ordinal* property unlike other common forms of positive dependence between random variables[11]. Given the comparative static result from Theorem 1, productivity, output, and capacity are increasing functions of s. Because s is positively associated by Theorem 2, productivity, output, and capacity will be positively associated as well (for fixed W).

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.

[10] Demand shocks are not correlated with cost side shocks.

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

2.3 Identification and input price variation

(a, q, k) is positively associated for fixed W^T. For all increasing functions ϕ_1 and ϕ_2,

$$\text{cov}\left[\phi_1\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right), \phi_2\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right) | w^T\right] \geq 0. \tag{17}$$

With data on (q, z, k), this assumption meaningfully identifies the production function and forms the basis for the bounds I construct in my application.

I still have one last problem to solve: we do not usually observe w^T. This is a problem because higher input prices tend to reduce output choice, but it is ambiguous whether higher input prices increase or decrease capacity and productivity choice. Higher variable input prices reduce the general scale of the plant which would reduce capacity and productivity choice. But higher input prices also encourage substitution to capacity and productivity by reducing the relative price of these factors of production. We do not know whether $a(\cdot, t)$ and $k(\cdot, t)$ are increasing or decreasing in input prices.

There are two ways to deal with this problem. The first way is to assume that input prices can be proxied by observables. The second way is to make the assumptions necessary to conclude that (a, q, z) are positively associated *unconditional* on input prices. I detail both approaches below, and use the second in my application.

The standard approach to deal with unobserved input price variation is to assume that w^T can be written as a function of observable plant characteristics x. The identification condition is then:

$$\text{cov}\left[\phi_1\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right), \phi_2\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right) | x\right] \geq 0. \tag{18}$$

While not ideal, this assumption is not any more restrictive than the status quo assumption employed in the literature when input prices are unobserved.

I develop a second approach that removes the need to observe or proxy input prices. The advantage of this approach is that it allows for an arbitrary underlying distribution of unobserved input prices if we believe the additional assumptions needed. The idea is to argue $s' = \left(s, -W^T\right)$ can play the role of s in the above results. Then, (a, q, k) is positively associated *unconditional* on input prices.

The starting point of the approach is to directly make the high-level assumption that the output-reduction effect of increasing input prices dominates the substitution effect. When variable input prices rise, productivity and capacity choice fall (Assumption 10). Many workhorse production models satisfy this assumption. For example, price-taking plants with Cobb Douglas production functions have input demand functions that are decreasing in the price of all inputs.

This approach also assumes variables inputs are *normal*. Normal inputs have increasing conditional input demand functions. The best argument for inputs being normal is the comparative static that would result if inputs were *not* normal. If inputs are not normal, then increases in input prices *increase* the amount of output plants produce. It is difficult to believe that increasing fuel prices, for example, would cause a power plant to generate more electricity. All inputs are normal in a Cobb-Douglas production function.

Assumption 10. $k\left(s, W^T, t\right)$ is *decreasing* in W^T (the net effect of an increase in input prices is for plants to weakly reduce their capacity) and $a\left(s, W^T, t\right)$ is *decreasing* in W^T (the net effect of an increase in input prices is for plants to weakly reduce their productivity).

Assumption 11. Inputs are *normal* (not inferior), they have increasing conditional input functions, or, by the envelope theorem,

$$\frac{\partial}{\partial Q} Z_\ell(Q, W) = \frac{\partial C}{\partial Q \partial W_\ell} \geq 0. \tag{19}$$

Next, define $s' = \left(s, -W^T\right)$ and define $s'_t = (s_t, -W_t)$. I assume the same Markov structure I put on s holds for s' as well. Theorem 2 then implies s' is positively associated, given Assumptions 12, 13, and 14.

Assumption 12. Define $s'_t = (K_0, A_0, \xi_t, \lambda_t, -W_t)$. State variables evolve according to the rule,

$$s'_{j,t} = \sigma'_{j,t}\left(s'_{t-1}, \upsilon'_{j,t}\right), \tag{20}$$

where $\sigma'_{j,t}$ is an increasing function for all (j, t) where j indexes the elements of s'_t.

Assumption 13. $(\xi_0, \lambda_0, A_0, K_0, -W_0)$ is positively associated.

Assumption 14. $\upsilon'_t = \left(\upsilon'_{1,t}, \ldots, \upsilon'_{J,t}\right)$ is a positively associated vector. Define $(s')^\tau = \cup_{t=0}^{\tau} s'_t$. $(s')^{t-1}$ and υ'_t are independent.

We then have Theorem 3. Under these assumptions, the same monotone comparative static results holds for s' as holds for s in Theorem 1.

Theorem 3. Given Assumptions 1, 2, 3, 4, 5, 6, 10, and 11, $a(s', t)$, $k(s', t)$, and $q(s', t)$ are increasing functions of $s' = \left(s, -W^T\right)$ for each t.

Proof. See Appendix B. □

$a(s', t)$, $k(s', t)$, and $q(s', t)$ are all increasing functions of s'. s' is positively associated given the same Markovian structure I put on s above. Therefore, (a, k, q) are positively associated unconditional on W^T. We have the following inequality restrictions,

$$\operatorname{cov}\left[\phi_1\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right), \phi_2\left(q_t - f\left(z_t, k_t\right), q_t, k_t\right)\right] \geq 0, \tag{21}$$

which do not depend on input prices. These inequalities are the main identification assumptions I use.

2.4 Increasing production function

I also assume the production function is increasing.

Assumption 15. The production function is increasing,

$$Df\left(z, k\right) \geq 0. \tag{22}$$

3 Inference on statistics of the productivity distribution

I use the identification results from Section 2 to make inference on parameters of the productivity distribution. I develop a set of assumptions that are implied by positive association and easier to use to construct bounds on parameters of interest in practice (Section 3.1 and 3.2). I then propose a Bayesian method of making inference on the partially-identified parameter of interest based on results from Kline and Tamer (2016) (Section 3.3).

Throughout, I focus on identifying coefficients in a regression of productivity on plant-level characteristics *not* on identifying the production function itself. While the bounds in Section 2 are less informative about individual output elasticities, they can put much narrower bounds on this statistic of the productivity distribution. To make the problem concrete, I consider estimating bounds on the parameter τ in the linear instrumental variable regression model,

$$a_t = x_t^\top \beta + u_t \tag{23}$$

$$\mathbb{E}\left[v_t u_t\right] = 0 \tag{24}$$

$$\tau = \rho^\top \beta, \tag{25}$$

where ρ is a known vector, x is a vector of observed covariates, and v is a vector of instruments. 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.

3.1 Linear positive association: a convex identified set

Positive association of (a, k, q) implies the following covariance restrictions,

$$\text{cov}\left[\phi_1(a, k, q), \phi_2(a, k, q)\right] \geq 0 \quad \text{for all increasing functions } \phi_1 \text{ and } \phi_2. \tag{26}$$

The set identified by positive association is the intersection of infinitely many covariance restrictions[12]. It is difficult to impose these restrictions directly to make inference on τ. To form bounds on τ, we would need to solve a complicated, nonconvex programming problem. I avoid this problem by introducing a weaker assumption than positive association which I call, "linear positive association". Linear positive association makes finding the bounds on τ a *convex* programming problem which is much easier to solve.

Suppose there is no plant that cannot produce any output no matter how many inputs it uses.

Assumption 16. There is a minimal level of productivity $(A \geq \underline{A} > 0)$. Intuitively, there exists no plant that can not produce any output no matter how many inputs it uses.

Given Assumption 16, it is a normalization to assume that $a \geq 0$.

Linear positive association identifies the log production function in the following set,

$$\mathcal{F}_{\text{LPA}} = \{f : Df \geq 0, q - f = a, a \geq 0, \tag{27}$$

$$\text{cov}\left[a\phi_1(q, k), \phi_2(q, k)\right] \geq 0 \tag{28}$$

$$\text{where } \phi_1 \geq 0 \text{ and } \phi_2 \text{ are increasing.}\}. \tag{29}$$

Linear positive association is weaker than positive association (given Assumption 16) because it uses a certain class of increasing functions of (a, k, q) instead of all increasing functions of (a, k, q). 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 4. 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,

$$\text{max or min }_{f \in \mathcal{F}_{\text{LPA}}} \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t f(z_t, k_t)\right] \tag{30}$$

[12] The bounds are intersection bounds in the sense of Chernozhukov, Lee, and Rosen (2013).

3.2 Computing the bounds via linear programming

Linear positive association simplifies computing the bounds even more in practice when we approximate the production function by a linear-in-parameters approximation like the Cobb Douglas or translog production function,

$$f(z,k) = r(z,k)^{\top}\theta, \tag{31}$$

where $r(z,k)$ is a known vector of functions and θ is an unknown parameter vector. The identified set for θ is the intersection of infinitely many linear inequalities,

$$\begin{aligned}\Theta_{\text{LPA}} = \{\theta : Dr\theta \geq 0, q - r(z,k)^{\top}\theta = a, a \geq 0,\\ \text{cov}[a\phi_1(q,k), \phi_2(q,k)] \geq 0 \quad \forall(\phi_1,\phi_2) \text{ st: } D\phi_1 \geq 0,\ D\phi_2 \geq 0\}.\end{aligned} \tag{32}$$

But because the production function is approximated parametrically it does not make sense to use all of the constraints included in Θ_{LPA}. If we were to, some of the constraints would have substantial identifying power via functional form[13]. For example, 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 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,

$$r(z,k) = (1,z,k). \tag{33}$$

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

$$\begin{aligned}\Theta_{\text{CD,LPA}} = \{\theta : \theta_z \geq 0, \theta_k \geq 0, q - \theta_1 - z^{\top}\theta_z - \theta_k k = a, a \geq 0,\\ \text{cov}[a\phi_1(q,k), \phi_2(q,k)] \geq 0,\\ \phi_1, \phi_2 \in \{\phi : \phi = \varphi_1 + \varphi_k k + \varphi_q q, \varphi_k \geq 0, \varphi_q \geq 0, \varphi_1 + \varphi_k k + \varphi_q q \geq 0\}\}\end{aligned} \tag{34}$$

The set $\Theta_{\text{CD,LPA}}$ can be constructed without considering infinitely many values of φ. Theorem 5 shows we only need to consider a finite set of values of φ. The identified set can be written as a finite number of linear inequalities.

[13] 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 5. 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,

$$\begin{aligned}\Theta_{\text{CD,LPA}} = \{\theta : & \theta_z \geq 0, \theta_k \geq 0, q - \theta_1 - z^\top \theta_z - \theta_k k = a, \\ & \underline{q} - \theta_1 - \overline{z}^\top \theta_z - \theta_k \overline{k} \geq 0, \\ & \text{cov}\left[a\phi_1(q,k), \phi_2(q,k)\right] \geq 0, \\ & \phi_1, \phi_2 \in \left\{\phi : \phi = \left(k - \underline{k}\right), \phi = \left(q - \underline{q}\right), \text{ or } \phi = 1\right\}\}\end{aligned} \tag{35}$$

Proof. See Appendix B. □

By Theorem 5, $\Theta_{\text{CD,LPA}}$ is a finite set of linear inequalities. The bounds on τ can be computed by solving two finite linear programming problems,

$$\text{max or min }_{\theta \in \Theta_{\text{CD,LPA}}} \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}} \mathbb{E}\left[v_t \left(1, z_t^\top, k_t\right)^\top\right]\theta. \tag{36}$$

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, μ (see (36)). The parameters μ are expectations of random variables which determine the covariance constraints in $\Theta_{\text{CD,LPA}}$ and the moments in the objective function: $\mathbb{E}\left[vx^\top\right]$, $\mathbb{E}\left[vy\right]$, and $\mathbb{E}\left[v\left(1, z^\top, k\right)^\top\right]$.

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 $\pi_n(\mu)$ 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, various 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 $\pi_n(\mu)$ converge in total variation as sample size increases,

$$\sup_h \left| \int h d\pi(\mu|\text{data}) - \int h d\pi_n(\mu) \right| \to 0, \tag{37}$$

independently of the prior specified for μ[14]. This result allows us to use the posterior distribution $\pi_n(\mu)$ as an approximation to $\pi(\mu|\text{data})$ 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,

$$\rho_n(\tau_0) = \int_\mu \mathbf{1}\{\tau_0 \in [\tau_{LB}(\mu), \tau_{UB}(\mu)]\} \, d\pi_n(\mu), \tag{38}$$

where the posterior of μ is approximated by π_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 will always depend on our prior no matter how much data we accumulate because τ is only partially identified by the data.

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. In addition, credible intervals 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,

$$HI_{\alpha,n} = \{\tau_0 : \rho_n(\tau_0) \geq \alpha\}. \tag{39}$$

So the 10% hypothesis interval is the set of parameter values (τ) that we believe, given the data, are more than 10% likely to be included in the bounds.

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

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.

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

$$\frac{1}{B}\sum_{b=1}^{B}\mathbf{1}\left(\tau_{LB}^{(b)} \leq \tau_0 \leq \tau_{UB}^{(b)}\right) \geq \alpha \tag{40}$$

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. $\tau_{LB}(\mu)$ 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[15].

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 How restructuring in the electricity industry affected productivity

Having established a method for computing bounds on τ, I now study whether the bounds on τ are narrow enough for us to learn anything about real empirical problems. Does linear positive association restrict τ enough to update our priors? I study how restructuring in the

[15] 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.

electricity industry affected power plant productivity using the approach developed in this article.

4.1 Industry and policy background

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[16]. If a utility lowered its costs, it would receive a lower price, giving 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 still 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 4.8). 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.

[16] In some states, rewards were not based on average cost but on some approximation of marginal cost, but the main point is that the price was not set by market forces. See Knittel (2002) for more on these other policies.

4.1.1 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 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)

4.2 Data

I use the same data as Fabrizio, Rose, and Wolfram (2007) for the years 1981 to 1999[17]. The original data is from the Energy Information Administration (EIA), the Federal Energy Regulatory Commission (FERC), and the Rural Utilities Service (RUS).

The Fabrizio, Rose, and Wolfram (2007) 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[18] and longer to build a coal plant.

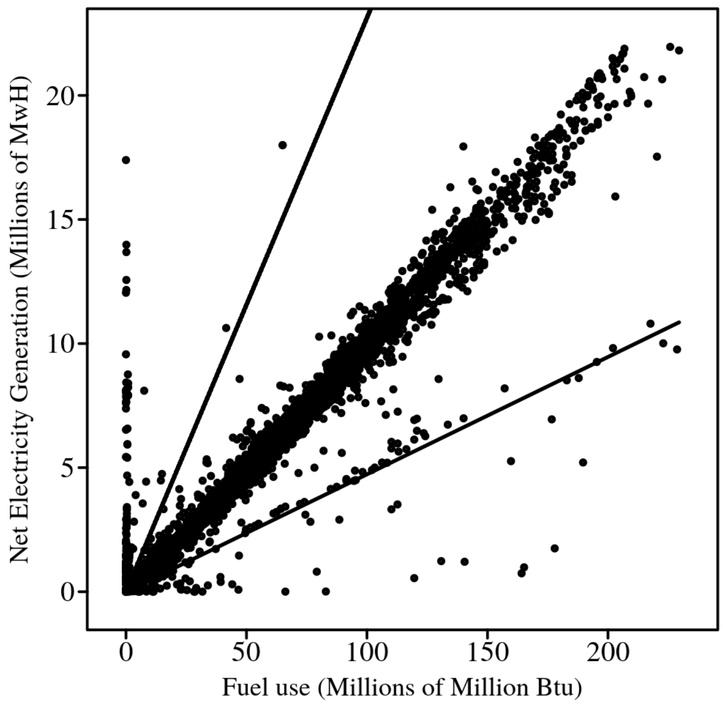
I measure output by millions of megawatt-hours of electricity generated net of electricity the power plant uses itself, fuel use by units of heat energy (in millions of million British thermal units), capacity by the total nameplate capacity of the plant (in megawatts), labor by total employment, and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

I include only power plants with positive employment and non-fuel expenses. Additionally,

[17] I thank Fabrizio, Rose, and Wolfram very much for making their data available and easy to use.

[18] 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



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.

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

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.

4.3 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,

$$Q^{(\sigma)} = \min\left\{H^{(\sigma)} F\left(L, E, K\right) A, K^{(\sigma)}\right\}. \tag{41}$$

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,

$$\sum Q^{(s)} = F\left(L, E, K\right) \sum_{s} H^{(s)} A \implies Q = F\left(L, E, K\right) HA, \tag{42}$$

because $H^{(s)} F\left(L, E, K\right) \leq K^{(s)}$ 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, K) as our vector of inputs,

$$\frac{Q}{H} = F\left(L, E, K\right) A. \tag{43}$$

Although $Q = H \times F(L, E, K)A$, 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,

$$\frac{Q}{H} = L^{\theta_\ell} E^{\theta_e} K^{\theta_k} A. \tag{44}$$

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

4.3.1 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 for 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. The two results do not contradict each other.

4.4 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.

4.4.1 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 $\mathrm{EverRestruct}_{it}$ 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,

$$x_{it} = \tau_X \times \text{Restruct}_{it} + \overline{\tau}_X Y_{it} \mathbf{1}\left(Y_{it} > 0\right) + \underline{\tau}_X Y_{it} + \kappa_X m_{rt} \tag{45}$$

$$+ \mu_{20} \times \mathbf{1}\left(\text{Age}_{it} \leq 20\right) + \mu_{30} \times \mathbf{1}\left(\text{Age}_{it} \leq 30\right) \tag{46}$$

$$+ \mu_{40} \times \mathbf{1}\left(\text{Age}_{it} \leq 40\right) + \iota \times \text{EverRestruct}_{it} + \xi_i + \delta_t + u_{it}, \tag{47}$$

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. 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.

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

The specification in Section 4.4.1 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:

$$x_{it} = \tau_X \times \text{Restruct}_{it}\text{IO}_{it} + \beta \times \text{Restruct}_{it} + \overline{\tau} Y_{it} \mathbf{1}\left(Y_{it} > 0\right) \text{IO}_{it} + \underline{\tau} Y_{it} \text{IO}_{it} \tag{48}$$
$$+ \overline{\beta} \times Y_{it} \mathbf{1}\left(Y_{it} > 0\right) + \underline{\beta} \times Y_{it} + \kappa m_{rt} \tag{49}$$
$$+ \mu_{20} \times \mathbf{1}\left(\text{Age}_{it} \leq 20\right) + \mu_{30} \times \mathbf{1}\left(\text{Age}_{it} \leq 30\right) + \mu_{40} \times \mathbf{1}\left(\text{Age}_{it} \leq 40\right) \tag{50}$$
$$+ \iota \times \text{EverRestruct}_{it} + \iota_{IOU} \times \text{EverRestruct}_{it}\text{IO}_{it} + \xi_i + \delta_t + u_{it}, \tag{51}$$

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.

4.5 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.

4.6 The effect of restructuring on power plant productivity

The effect of restructuring on fuel productivity is,

$$\tau_a = \tau_{q-h} - \tau_{f(\ell,e,k)}, \tag{52}$$

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%)

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 $\tau_{f(\ell,e,k)}$ 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 $f(\ell, e, k)$ 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,

$$\tau_f = \theta_\ell \tau_\ell + \theta_e \tau_e + \theta_k \tau_k, \tag{53}$$

where τ_ℓ, τ_e, and τ_k are estimates of the effect of restructuring on log labor, log nonfuel expenditures, and log capacity. 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, τ_e, and τ_k in Section 4.5 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 $(\theta_\ell, \theta_e, \theta_k) \geq 0$, if restructuring has a negative effect on input use ($(\tau_\ell, \tau_e, \tau_k) \leq 0$) as we saw in Section 4.5, then restructuring will have a negative effect on the value of the log production function f ($\tau_f = \theta_\ell \tau_\ell + \theta_e \tau_e + \theta_k \tau_k \leq 0$). So:

$$\tau_a = \tau_{q-h} - \tau_f \geq \tau_{q-h}. \tag{54}$$

Because $\tau_{q-h} < 0$ (see Section 4.5), the question is whether $(\theta_\ell, \theta_e, \theta_k)$ can be large enough that $\tau_a = \tau_{q-h} - \tau_\ell\theta_\ell - \tau_e\theta_e - \tau_k\theta_k$ can be positive or whether the linear positive association assumption sufficiently restricts how large θ can be so that $\tau_f = \theta_\ell \tau_\ell + \theta_e \tau_e + \theta_k \tau_k$ is not *that* negative[19]. 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 -4.75% and -0.45% across the two control group specifications using the 10% hypothesis intervals. Point estimates of the bounds are between -2.87% and -1.16%. An effect magnitude between 0.45% and 4.75% is economically meaningful: to maintain the same output as before, the plant would need to burn *at least* 0.45% to 4.75% more fuel if it has the same capacity, labor, and non-fuel expenditures and no capacity constraints bind.

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.

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

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.

To investigate whether the second case is relevant, I study the slope of the post-restructuring trend ($\overline{\tau}$). We should be cautious in interpreting the trend as 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. I estimate the annual trend in the effect of restructuring to be between 0.25% and 1.04% using investor-owned power plants as the control and between -0.64% and 0.86% 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 $\overline{\tau} \geq 0$),

$$\eta = -1 \times \frac{\tau}{\overline{\tau}}. \tag{55}$$

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 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[20].

The lower bound on η using Control Group 1 is 1.73 while the lower bound using Control Group 2 is 6.83. Because the maximum number of years post-restructuring in sample is 7, the results for Control Group 2 suggest that the productivity effect of restructuring is long-lasting. The results using Control Group 1 suggest the sign may flip relatively soon if we interpret $\overline{\tau}$ entirely causally. So the data might admit the disruption theory, but there is not strong evidence in favor of it.

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)[21].

[20] 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.

[21] 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.

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 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.

4.7 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

Table 3: Effect of restructuring on plant fuel productivity

Parameter	Control Group	LB (10%)	LB	UB	UB (10%)
τ_a	MUNI	−4.75%	−2.87%	−2.39%	−0.74%
	IO	−1.99%	−1.44%	−1.16%	−0.45%
κ_a	MUNI	0.80%	2.33%	3.25%	4.59%
	IO	0.88%	2.31%	3.13%	4.49%
$\overline{\tau}_a$	MUNI	−0.64%	−0.22%	0.35%	0.86%
	IO	0.25%	0.54%	0.68%	1.04%
$\underline{\tau}_a$	MUNI	0.10%	0.22%	0.36%	0.50%
	IO	−0.08%	−0.04%	0.08%	0.13%
β	MUNI	−0.48%	0.79%	1.45%	3.13%
$\overline{\beta}$	MUNI	−0.38%	0.27%	0.77%	1.36%
$\underline{\beta}$	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).

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,

$$o_{it} = \frac{Q_{it}}{\sum_{j \in \text{state}} Q_{jt}} \tag{56}$$

$$\text{GAP}_{\text{state},t} = \prod_{i \in \text{state}} A_{it}^{o_{it}}. \tag{57}$$

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:

$$\log \text{GAP} = \underbrace{\frac{1}{N}\sum_i a_i}_{\text{Average Productivity}} + \underbrace{\sum_i \left(a_i - \frac{1}{N}\sum_i a_i\right) \times \left(o_i - \frac{1}{N}\sum_i o_i\right)}_{\text{Reallocation term}}. \tag{58}$$

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,

$$x_{\text{state},t} = \tau_x \times \text{Restruct}_{\text{state},t} + \underline{\tau} Y_{\text{state},t} + \overline{\tau} Y_{\text{state},t} \times \mathbf{1}\left(Y_{\text{state},t} > 0\right) + \text{Year fixed effect} \tag{59}$$

$$+ \text{State effect} + u_{\text{state},t}, \tag{60}$$

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 -5.35% and -1.99% for investor-owned utilities and by -5.64% and -1.14% 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 0.78% and 3.05%. 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 -0.12% to 2.54%.

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.

4.8 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

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

Variable	Parameter	Plants Included	LB (10%)	LB	UB	UB (10%)
GAP	τ	IO	−3.74%	−2.17%	−1.44%	0.36%
		All	−4.28%	−2.29%	−1.78%	0.00%
	$\overline{\tau}$	IO	−0.30%	0.31%	0.91%	1.59%
		All	0.11%	0.70%	1.04%	1.61%
	$\underline{\tau}$	IO	−0.13%	−0.03%	0.06%	0.16%
		All	−0.10%	−0.03%	0.09%	0.19%
Average Productivity	τ	IO	−5.35%	−3.93%	−3.43%	−1.99%
		All	−5.64%	−3.81%	−2.69%	−1.14%
	$\overline{\tau}$	IO	−0.52%	0.08%	1.08%	1.61%
		All	−0.27%	0.34%	0.79%	1.43%
	$\underline{\tau}$	IO	0.01%	0.10%	0.22%	0.32%
		All	−0.00%	0.09%	0.20%	0.32%
Reallocation	τ	IO	0.78%	1.70%	2.14%	3.05%
		All	−0.12%	0.85%	1.53%	2.54%
	$\overline{\tau}$	IO	−0.65%	−0.38%	0.23%	0.59%
		All	−0.26%	0.06%	0.36%	0.72%
	$\underline{\tau}$	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.

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.

5 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.

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). Identifying market power from production data. *Working paper*.

Gandhi, A., S. Navarro, and D. Rivers (2015). 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).

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*.

Jones, C. (2005). The shape of production functions and the direction of technical change. *Quarterly Journal of Economics 120*(2), 517–549.

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

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.

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.

Uras, B. and P. Wang (2017). Production flexibility, misallocation and total factor productivity. *NBER Working Paper* (23970).

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.

Wolak, F. (1994). An econometric analysis of the asymmetric information, regulator-utility interaction. *Annales d'Économie et de Statistique* (34), 13–69.

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

$$\mathcal{F}_{\text{LPA}} = \{f : Df \geq 0, q - f = a, a \geq 0, \tag{61}$$

$$\operatorname{cov}\left[a\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq \left(u_{1,q}, u_{1,k}\right)\right],\right. \tag{62}$$

$$\left.\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq \left(u_{2,q}, u_{2,k}\right)\right]\right] \geq 0 \quad \forall \left(u_{1,q}, u_{1,k}, u_{2,q}, u_{2,k}\right) \in [0,1]^4\} \tag{63}$$

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

Proof. See Appendix B. $\square$

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,

$$r_{p,m}(z,k) = \mathbf{1}\left(\left(\{\varphi_{z_\ell}(z_\ell)\}_{\ell=1}^{L}, \varphi_k(k)\right) \geq \left(\left\{\frac{p_{z_\ell}}{m}\right\}_{\ell=1}^{L}, \frac{p_k}{m}\right)\right), \tag{64}$$

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).

$$\begin{aligned}\Theta_{m_1,m_2,\text{LPA}} = \{\theta : \theta_{-0} \geq 0, q - r(z,k)^\top \theta = a, a \geq 0,& \\ \operatorname{cov}\left[a\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq \left(u_{1,q}, u_{1,k}\right)\right],\right.& \\ \left.\mathbf{1}\left[\left(\varphi_q(q), \varphi_k(k)\right) \geq \left(u_{2,q}, u_{2,k}\right)\right]\right] \geq 0& \\ \forall \left(u_{1,q}, u_{1,k}\right) \in \left\{0, \frac{1}{m_1}, \ldots, \frac{m_1 - 1}{m_1}\right\}^2 \text{ and}& \\ \forall \left(u_{2,q}, u_{2,k}\right) \in \left\{0, \frac{1}{m_2} \ldots, \frac{m_2 - 1}{m_2}\right\}^2\}&\end{aligned} \tag{65}$$

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,

$$\text{max or min }_{\theta\in\Theta_{m_1,m_2,\text{LPA}}} \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}}\mathbb{E}\left[v_t q_t\right] - \mathbb{E}\left[vx^\top\right]^{-1}_{\text{left}}\mathbb{E}\left[v_t r\left(z_t, k_t\right)^\top\right]\theta. \tag{66}$$

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 Theorem 1

Proof of Theorem 1.

Consider the following optimization problem (where A_0, K_0 are given),

$$\max_{K_t,Q_t,A_t} \quad \sum_{t=1}^{T}\beta_t \times \left[P\left(Q_t,\xi_t\right)Q_t - C\left(\frac{Q_t}{A_t}, K_t, W_t\right) - M\left(A_t, A_{t-1}, \lambda_t\right) - G\left(K_{t+1}, K_t\right)\right] \tag{67}$$

$$\text{st:} \quad 0 \le Q_t \le K_t,\ A_t \ge 0 \tag{68}$$

I use the monotone comparative statics theorem from Topkis (1978) to establish Theorem 1. The Topkis (1978) theorem requires that the objective function is supermodular in $(Q_t, K_t, A_t)_{t=1}^T$ and has increasing differences in the choices $(Q_t, K_t, A_t)_{t=1}^T$ and the parameters $\left(A_0, K_0, (\xi_t, \lambda_t)_{t=1}^T\right)$. Supermodularity can be established by showing all pairwise cross partials of the choices are positive and increasing differences can be established by showing the cross partial of each choice variable with each parameter is positive. I check all of these derivatives and show the above assumptions imply they are all positive.

The other requirement of the Topkis (1978) theorem is that the set over which we are optimizing is a lattice. That is, every two elements from the set have their "meet" (in this context, the elementwise minimum of the two vectors) included in the set and their "join" (in this context, the elementwise maximum of the two vectors) included in the set.

The constraint set is a lattice because if Q and Q' are both nonnegative and less than or equal to capacities K and K', then, $0 \le \min\{Q, Q'\} \le \min\{K, K'\}$ and $0 \le \max\{Q, Q'\} \le \max\{K, K'\}$ so if $(Q_t, K_t, A_t)_{t=1}^T$ and $(Q'_t, K'_t, A'_t)_{t=1}^T$ are two feasible choices, both the meet and join of the two are feasible.

I organize the proof the objective function is supermodular and has increasing difference in choices and parameters into three sections, one for each kind of choice variable (Q_t, A_t, K_t), first stating what the derivative is with respect to the choice variable and then showing that derivative is increasing in all parameters and other choice variables.

- The derivative with respect to A_t (define $\beta_{T+1} = 0$ when considering the $t = T$ case) is,

$$\beta_t \times \left[C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t^2} - \frac{\partial M}{\partial A_t}(A_t, A_{t-1}, \lambda_t) \right] \tag{69}$$

$$-\beta_{t+1} \times \frac{\partial M}{\partial A_t}(A_{t+1}, A_t, \lambda_{t+1}) \tag{70}$$

I now go through each parameter or choice in the above expression to show the derivative is increasing in it.

The cross partial with respect to (A_t, Q_t) is,

$$\beta_t \times C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t^3} + \frac{1}{A_t^2} \times C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \tag{71}$$

$$= \frac{\beta_t}{A_t^2} \times \left[C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t} + C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \right]. \tag{72}$$

By Assumption 4 and because marginal cost is positive,

$$C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{\frac{Q_t}{A_t}}{C'\left(\frac{Q_t}{A_t}, K_t, W_t\right)} \geq -1 \tag{73}$$

$$\iff C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t} \geq -C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \tag{74}$$

$$\iff C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t} + C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \geq 0 \tag{75}$$

So, the cross partial with respect to (A_t, Q_t) is nonnegative,

$$\frac{\beta_t}{A_t^2} \times \left[C''\left(\frac{Q_t}{A_t}, K_t, W_t\right) \times \frac{Q_t}{A_t} + C'\left(\frac{Q_t}{A_t}, K_t, W_t\right) \right] \geq 0. \tag{76}$$

The cross partial with respect to (A_t, A_{t+1}) is,

$$-\beta_{t+1} \frac{\partial M}{\partial A_t \partial A_{t+1}} \geq 0 \tag{77}$$

where the inequality holds by Assumption 6 (the cross partial of M is negative).

The cross partial with respect to (A_t, A_{t-1}) is,

$$-\beta_t \times \frac{\partial M}{\partial A_t \partial A_{t-1}} \geq 0, \tag{78}$$

where the inequality holds again by Assumption 6.

The two plant parameters that appear in the derivative with respect to A_t are λ_t and λ_{t+1}. The cross partial with respect to (A_t, λ_t), and the cross partial with respect to (A_t, λ_{t+1}) are both positive by Assumption 2.

$$(A_t, \lambda_t): \quad -\beta_t \times \frac{\partial M}{\partial \lambda \partial A_t}(A_t, A_{t-1}, \lambda_t) \geq 0 \tag{79}$$

$$(A_t, \lambda_{t+1}): \quad -\beta_{t+1} \frac{\partial M}{\partial A_t \partial \lambda}(A_{t+1}, A_t, \lambda_{t+1}) \geq 0 \tag{80}$$

The rest of the parameters and choices do not appear in the derivative with respect to A_t so the rest of the cross-partials are 0, and so, greater than or equal to 0.

- The derivative with respect to Q_t is,

$$\beta_t \times \left[\frac{\partial P}{\partial Q}(Q_t, \xi_t) Q_t + P(Q_t, \xi_t) - \frac{\partial C}{\partial \widetilde{Q}}\left(\widetilde{Q}_t, K_t, W_t\right) \times \frac{1}{A_t}\right] \tag{81}$$

As shown above, the cross partial (Q_t, A_t) is nonnegative.

The cross partial (Q_t, K_t) is,

$$-\beta_t \frac{\partial C}{\partial \widetilde{Q} \partial K}\left(\widetilde{Q}_t, K_t, W_t\right) \times \frac{1}{A_t} \geq 0, \tag{82}$$

where the inequality holds because marginal variable cost weakly decreases in capacity by Assumption 3.

The cross partial (Q_t, ξ_t) is,

$$\beta_t \times \frac{\partial}{\partial \xi_t}\left[\frac{\partial P}{\partial Q}(Q_t, \xi_t) Q_t + P(Q_t, \xi_t)\right] = \beta_t \times \frac{\partial}{\partial \xi_t} \mathrm{MR}(Q_t, \xi_t) \geq 0, \tag{83}$$

by Assumption 1.

The rest of the parameters and choices do not appear in the derivative with respect to Q_t so the rest of the cross-partials are 0, and so, greater than or equal to 0.

- The derivative with respect to K_t is (for $t \leq T$, $K_{T+1} = 0$ always),

$$-\beta_t \times \left[\frac{\partial C}{\partial K}\left(\widetilde{Q}_t, K_t, W_t\right) + \frac{\partial G}{\partial K_t}(K_{t+1}, K_t)\right] - \beta_{t-1} \frac{\partial G}{\partial K_t}(K_t, K_{t-1}) \tag{84}$$

The cross partial with respect to (K_t, Q_t) is nonnegative (see above).

The cross partial with respect to (K_t, K_{t+1}) is,

$$-\beta_t \times \frac{\partial G}{\partial K_{t+1} \partial K_t}(K_{t+1}, K_t) \geq 0, \tag{85}$$

where the inequality holds by Assumption 5.

The cross partial with respect to (K_t, K_{t-1}) is,

$$-\beta_{t-1} \frac{\partial G}{\partial K_t \partial K_{t-1}} (K_t, K_{t-1}) \geq 0, \tag{86}$$

where the inequality holds by Assumption 5.

The rest of the parameters and choices do not appear in the derivative with respect to K_t so the rest of the cross-partials are 0, and so, greater than or equal to 0.

Because the objective function is supermodular in $(Q_t, K_t, A_t)_{t=1}^T$ and has increasing differences in $(Q_t, K_t, A_t)_{t=1}^T$ and $s = \left(A_0, K_0, \{\lambda_t\}_{t=1}^T, \{\xi_t\}_{t=1}^T\right)$, by the Topkis (1978) theorem, the choice variables $\left(\widetilde{Q}_t, K_t, A_t\right)_{t=1}^T$ are increasing functions of s, holding the non-s parameters of the optimization problem fixed, establishing the conclusion of the theorem.

□

B.2 Proof of Theorem 2

Proof of Theorem 2.

Recall that $s = \cup_{t=0}^T s_t$. By Assumption 8, s^0 is positively associated.

By way of induction, suppose that s^t is positively associated. By Assumption 7,

$$s_{t+1} = \sigma_t (s_t, \upsilon_{t+1}). \tag{87}$$

Because υ_{t+1} is positively associated and independent of s^t, (s^t, υ_{t+1}) are positively associated. Because σ_t is an increasing function of (s^t, υ_{t+1}), $(s^t, \upsilon_{t+1}, \sigma_t(s_t, \upsilon_{t+1}))$ is positively associated as well, implying that $(s^t, s_{t+1}) = s^{t+1}$ is positively associated.

By induction, we have that s^τ is positively associated for all τ including, in particular, $\tau = T$. So $s = s^T$ is positively associated. □

B.3 Proof of Theorem 3

Proof of Theorem 3.

Consider the interior optimization problem for output for a fixed choice of (K_t, A_t):

$$\max_Q \sum_{t=1}^T \beta_t \times \left[P(Q_t, \xi_t) Q_t - C\left(\frac{Q_t}{A_t}, K_t, W_t\right) - M(A_t, A_{t-1}, \lambda_t) - G(K_{t+1}, K_t)\right] \tag{88}$$

Because Q_t has increasing differences with K_t and A_t and the rest of the parameters of the problem aside from input prices as shown in Theorem 1 and by Assumption 11 Q_t has decreasing differences with W_t,

$$-\frac{\partial \widetilde{Q} \partial W}{C}\left(\widetilde{Q}_t, K_t, W_t\right) \leq 0, \tag{89}$$

and so increasing differences with $-W_t$, by the Topkis (1978) theorem

$$Q_t = Q_t\left(K_t, A_t, -W_t, s\right)$$

is increasing in all arguments.

Because $K_t\left(s, -W^T\right)$ and $A_t\left(s, -W^T\right)$ are increasing functions of both s (by Theorem 1) and $-W^T$ (by Assumption 10), Q_t is also an increasing of $-W^T$ and s because the total derivative of Q_t with respect to W_t is,

$$\frac{\partial Q_t}{\partial K_t} \times \frac{\partial K_t}{\partial W_t} + \frac{\partial Q_t}{\partial A_t} \times \frac{\partial A_t}{\partial W_t} - \frac{\partial Q_t}{\partial - W_T} \leq 0, \tag{90}$$

so the derivative of Q_t with respect to $-W_t$ is positive because Q_t is increasing in $(K_t, A_t, -W_t)$.

Therefore: $a\left(s, -W^T, t\right)$, $k\left(s, -W^T, t\right)$, and $q\left(s, -W^T, t\right)$ are increasing functions of $\left(s, -W^T\right)$ for each t. $\square$

B.4 Proof of Theorem 4

Proof of Theorem 4.

The derivative of f^α is,

$$Df^\alpha = (1-\alpha)\, Df^0 + \alpha Df^1 \geq 0, \tag{91}$$

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

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

$$\operatorname{cov}\left[\left(q - f^0\right)\phi_1, \phi_2\right] \geq 0, \quad \operatorname{cov}\left[\left(q - f^1\right)\phi_1, \phi_2\right] \geq 0 \tag{92}$$

$$\implies \alpha \operatorname{cov}\left[\left(q - f^0\right)\phi_1, \phi_2\right] + (1-\alpha)\operatorname{cov}\left[\left(q - f^1\right)\phi_1, \phi_2\right] \geq 0 \tag{93}$$

$$\implies \operatorname{cov}\left[\left(q - f^\alpha\right)\phi_1, \phi_2\right] \geq 0 \tag{94}$$

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

B.5 Proof of Theorem 5

Proof of Theorem 5.

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:

$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(q - \underline{q}\right), \left(q - \underline{q}\right)\right) \geq 0 \tag{95}$$
$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(q - \underline{q}\right), \left(k - \underline{k}\right)\right) \geq 0 \tag{96}$$
$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(k - \underline{k}\right), \left(k - \underline{k}\right)\right) \geq 0 \tag{97}$$
$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(k - \underline{k}\right), \left(q - \underline{q}\right)\right) \geq 0 \tag{98}$$

We want to show that θ satisfies,

$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left[\varphi_{1,1} + \varphi_{1,k} k + \varphi_{1,q} q\right], \varphi_{2,1} + \varphi_{2,k} k + \varphi_{2,q} q\right) \geq 0 \tag{99}$$

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 $\varphi_{2,q}\left(q - \underline{q}\right)$,

$$\varphi_{1,q}\varphi_{2,q}\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(q - \underline{q}\right), \left(q - \underline{q}\right)\right) \tag{100}$$
$$= \text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\varphi_{1,q}\left(q - \underline{q}\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \geq 0 \tag{101}$$

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

$$\varphi_{1,k}\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(k - \underline{k}\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \tag{102}$$
$$= \text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\varphi_{1,k}\left(k - \underline{k}\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \geq 0 \tag{103}$$

- 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 $\varphi_{2,q}\left(q - \underline{q}\right)$,

$$\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\varphi_{1,k}\left(k - \underline{k}\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \tag{104}$$
$$+\text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\varphi_{1,q}\left(q - \underline{q}\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \tag{105}$$
$$= \text{cov}\left(\left[q - \theta_1 - z^\top \theta_z - \theta_k k\right]\left(\varphi_{1,k}\left(k - \underline{k}\right) + \varphi_{1,q}\left(q - \underline{q}\right)\right), \varphi_{2,q}\left(q - \underline{q}\right)\right) \geq 0 \tag{106}$$

- 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 $\varphi_{2,q}\left(q-\underline{q}\right)+\varphi_{2,k}\left(k-\underline{k}\right)$,

$$\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\varphi_{2,q}\left(q-\underline{q}\right)\right) \tag{107}$$
$$+\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\varphi_{2,k}\left(k-\underline{k}\right)\right) \tag{108}$$
$$=\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\right. \tag{109}$$
$$\left.\varphi_{2,q}\left(q-\underline{q}\right)+\varphi_{2,k}\left(k-\underline{k}\right)\right)\geq 0 \tag{110}$$

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 $\operatorname{cov}\left(X_1,X_2+\text{constant}\right)=\operatorname{cov}\left(X_1,X_2\right)$. 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,

$$\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\Delta,\right. \tag{111}$$
$$\left.\varphi_{2,1}+\varphi_{2,q}q+\varphi_{2,k}k\right) \tag{112}$$
$$+\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\left(\varphi_{1,k}\left(k-\underline{k}\right)+\varphi_{1,q}\left(q-\underline{q}\right)\right),\right. \tag{113}$$
$$\left.\varphi_{2,1}+\varphi_{2,q}q+\varphi_{2,k}k\right). \tag{114}$$

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

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,

$$\operatorname{cov}\left(\left[q-\theta_1-z^\top\theta_z-\theta_k k\right]\left(\varphi_{1,1}+\varphi_{1,k}k+\varphi_{1,q}q\right),\right. \tag{115}$$
$$\left.\varphi_{2,1}+\varphi_{2,q}q+\varphi_{2,k}k\right)\geq 0. \tag{116}$$

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.6 Proof of Theorem 6

Proof of Theorem 6.

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,

$$\eta_i = \varphi_i^{-1}\left(\max_{j \in \{1,\ldots,k\}} \{t_j : \mathbf{1}(\varphi_i(z_i) \geq t_j) = 1\}\right) \text{ for } i = 1, \ldots, L. \tag{117}$$

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

Assume,

$$\operatorname{cov}(a\mathbf{1}(\varphi(z_i) \geq u), \mathbf{1}(\varphi(z_i) \geq v)) \geq 0. \tag{118}$$

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

There is a positive function κ_h such that,

$$h_k(z) = \sum_{t_1=1}^{k} \cdots \sum_{t_L=1}^{k} \kappa_h(t_1, \ldots, t_L) \times \mathbf{1}(\varphi_1(z_1) \geq t_1, \ldots, \varphi_L(z_L) \geq t_L). \tag{119}$$

There is such a κ_v for v_k as well.

Therefore, by the linearity of the covariance,

$$\sum_{t^1} \sum_{t^2} \kappa_h\left(t^1\right) \kappa_v\left(t^2\right) \operatorname{cov}\left(a\mathbf{1}\left(\varphi(z) \geq t^1\right), \mathbf{1}\left(\varphi(z) \geq t^2\right)\right) = \tag{120}$$

$$\operatorname{cov}(ah_k(z), v_k(z)) \geq 0. \tag{121}$$

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 $\{t_1, \ldots, t_k\} \to \mathbb{Q} \cap [0, 1]$). Because ϕ_1 and ϕ_2 are bounded and because the sequences in k are increasing, by the monotone convergence theorem,

$$\mathbb{E}(ah_k(z)) \to \mathbb{E}(a\phi_1(z)) \tag{122}$$

$$\mathbb{E}(v_k(z)) \to \mathbb{E}(\phi_2(z)) \tag{123}$$

$$\mathbb{E}(ah_k(z)\, v_k(z)) \to \mathbb{E}(a\phi_1(z)\, \phi_2(z)). \tag{124}$$

So I have what I wanted to show,

$$0 \leq \lim_{k\to\infty} \operatorname{cov}(ah_k(z), v_k(z)) = \operatorname{cov}(a\phi_1(z), \phi_2(z)). \tag{125}$$

□