

Identifying productivity when it is a factor of production

Zach Flynn*

August 19, 2019

Abstract

Economists typically model a plant's productivity as an exogenous characteristic, but the people who run and work at manufacturing plants make choices, at a cost, that affect plant productivity. I develop a method to partially identify the productivity distribution when such choices determine productivity. The method uses a monotone comparative static result I prove in a general economic model. It does not require instruments or timing assumptions. I use the method to study the effect of implementing market-based pricing on productivity in the electricity generation industry.

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

1 Introduction

In empirical work, economists typically model productivity, the residual to the production function, as an exogenous shock plants react to¹. But plants make choices, at a cost, 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. The choices of *people* in response to the incentives they face determine whether an organization is productive. This fact motivates the development of a general empirical strategy to deal with endogenous productivity.

I model productivity as an unobserved factor of production. Under general conditions, I prove a comparative static result: plant productivity, output, and capacity choices increase in the latent, unobserved variables that determine the plant’s residual demand curve and its various costs. I show that when these unobserved state variables follow a standard Markov process, productivity, capacity, and output are *positively associated* in the sense of [Esary, Proschan, and Walkup \(1967\)](#): the covariance of any two increasing functions of productivity, output, and capacity is positive. While productivity itself is unobservable, with data on output, capacity, and input use, we can construct the set of production functions such that the implied productivity is positively associated with output and capacity. Doing so partially identifies the production function because, given the data, some ex-ante reasonable production functions imply productivity distributions that do not satisfy the restriction. I develop the model, comparative static, and identification result in [Section 2](#).

I then show how to use this identification result to make inference on coefficients in regressions where log productivity is the dependent variable ([Section 3](#)). Such coefficients are a standard statistic of interest in the productivity literature. We use them when we want to evaluate the effect of a policy on productivity² or to learn which kinds of plants are more productive than others. I show the bounds on these coefficients are the values of two easy-to-compute linear programming problems. To make inference on these coefficients, I use a Bayesian strategy based on results from [Kline and Tamer \(2016\)](#). I compute the posterior distribution of the bounds via simulation.

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

In [Section 5](#), I demonstrate the bounds are narrow enough to be useful in practice by using them in practice. I bound the effect of introducing market-based pricing (“restructuring”) in the electricity generation industry on power plant productivity. Historically, state-run public service commissions set electricity prices on the basis of the costs the utilities incurred in

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

²For example, [Pavcnik \(2002\)](#) uses such regressions to understand how trade liberalization in Chile affected productivity.

producing the electricity. In the mid 1990’s to early 2000’s, some US states restructured the industry by allowing markets to set electricity prices instead of public service commissions. Other states maintained regulated pricing. When states set prices on the basis of costs, plants have less of an incentive to reduce those costs so ending regulated pricing might have encouraged power plants to choose to be more productive. But restructuring has other effects that encourage power plants to reduce their productivity. For example, restructuring forced utilities to disintegrate so that transmission, the naturally monopolistic stage of electricity production, could be regulated while the state allowed markets to set prices in the generation stage. Any efficiencies from integration were lost with restructuring. The problem is a good example of the empirical relevance of allowing for endogenous productivity because the policy might affect plant productivity by changing the plants’ incentives.

I study whether power plants increased their “capacity-agnostic productivity” because of the restructuring policy. A power plant has a higher capacity-agnostic productivity if it produces more output, holding all inputs constant, whenever it is not capacity-constrained. This multi-factor productivity measure appears in a natural model of electricity production I develop in Section 5.3. I find restructuring caused power plants to lower their capacity-agnostic productivity by between 1.12% and 2.87%. Because productivity is a choice plants make at a cost, if a policy lowers productivity, it does not necessarily reduce welfare.

This article primarily contributes to the literature on identifying the production function and measuring productivity. The basic problem of identifying the production function is that input choice is correlated with productivity. This problem applies whether productivity is endogenously chosen or it is an exogenous parameter of the plant’s production function. A large and old literature discusses this identification problem 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 \(2003\)](#), [Akerberg, Caves, and Frazer \(2015\)](#), and [Gandhi, Navarro, and Rivers \(2019\)](#) based on the models of firm and industry dynamics developed in [Jovanovic \(1982\)](#) and [Hopenhayn \(1992\)](#)³. Identification in the proxy approach is based on information and timing restrictions: restrictions on *what* the plant knows *when*. I add to this literature a new approach not based on such restrictions, but instead based on comparative statics derived from basic economic restrictions on the primitives of the plant’s problem.

[Doraszelski and Jaumandreu \(2013\)](#) and [De Loecker \(2013\)](#) modify the standard, exogenous productivity proxy model to develop the primary method in the prior literature to estimate endogenous productivity models. They allow plants to control productivity with *observed* choices⁴. They assume these choices are made before any unobserved shock to productivity is revealed to the plant. Relative to this work, I allow for unobserved investment in productivity and do not assume when the plant makes the investment. I also do not assume productivity

³The proxy approach has been applied to a wide-range of empirical problems. These applications include studying the effects of trade liberalization ([Pavcnik, 2002](#)) and the effect of restructuring on the telecommunications equipment industry ([Olley and Pakes, 1996](#)) and, more recently, to estimate markups ([De Loecker and Warzynski, 2012](#); [De Loecker, Eeckhout, and Unger, 2019](#); [Flynn, Gandhi, and Traina, 2019](#)). See [Syverson \(2011\)](#) for a broader survey of applications.

⁴[Van Biesebroek \(2003\)](#) also studied observed technology choice and how it affected productivity where he observes automobile plants adopting different technologies.

is the only unobserved state variable differentiating plants or that productivity is a function of observables, the two calling cards of the proxy approach to production function estimation.

This article also contributes to the literature studying the effect of restructuring on power plant productivity. Fabrizio, Rose, and Wolfram (2007) is the most closely related analysis. They estimate the effect of restructuring on conditional input demand equations and use a proxy for demand (total electricity sales in a state, a measure of market size) to instrument for output. But demand is only potentially uncorrelated with *exogenous* productivity. If productivity is a factor of production, demand affects it like it affects the choice of any other factor of production. So their model implicitly assumes productivity is exogenous.

2 Partial identification in a general model of production via monotone comparative statics

I develop an identification strategy based on a monotone comparative static result that works when plants choose their productivity. The model I use to establish this result is general and encapsulates many other common models. The result holds when productivity is a static decision⁵ and when plants choose it with dynamic considerations⁶, when plants are capacity-constrained and when they are not constrained, when competition is perfect and when it is imperfect, and when productivity is a choice and when it is exogenous.

To give an outline of the identification strategy: I first show via monotone comparative static methods that output, capacity, and productivity choice are increasing in the unobserved state variables that determine a plant's cost function and its (residual) demand curve (Section 2.1). I pair this result with a Markovian model of how the unobserved state variables vary across plants (Section 2.2). This pairing leads to the result that productivity is *positively associated* with output and capacity in the sense of Esary, Proschan, and Walkup (1967): the covariance of any two increasing functions of output, capacity, and productivity is positive. I use this result to partially identify the production function and the productivity distribution (Section 2.3).

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

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

⁵For 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.

⁶For example, when productivity is a stock of capability like better machines or people.

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

$$\begin{aligned} \mathcal{F} = \{f \in \text{increasing functions} : & \text{for all functions } \phi_1 \text{ and } \phi_2 \\ & \text{that are increasing in all arguments ,} \\ & \text{cov} [\phi_1 (q - f (z, k) , q, k) , \phi_2 (q - f (z, k) , q, k)] \geq 0\} . \end{aligned} \quad (2)$$

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

2.1 The model and comparative static result

I describe each element of the model I use to derive this identification result and then introduce the profit maximization problem the plant solves:

- $P_t(Q, \xi_t)$ gives the (residual) demand function for each plant in period t . ξ indexes how the (residual) demand function varies across plants.
- $C_t(Q, A, K, W)$ is the cost of the Z inputs. Let Z be the variable inputs and W be input prices (which may vary by plant). When the plant is capacity-constrained, the cost function is,

$$C_t(Q, A, K, W) = \min W^\top Z \quad \text{ST:} \quad F(Z, K) \geq \frac{Q}{A} \text{ if } Q \leq K \quad (3)$$

$$= \infty \text{ if } Q > K. \quad (4)$$

I prove two comparative static results: one for when K is capital and the plant is not capacity-constrained and another for when K is capacity. When the plant is not capacity constrained, the cost function C is as above without the constraint that $Q \leq K$.

- $M_t(A_t, A_{t-1}, \eta_t)$ gives the cost of choosing a certain level of productivity, given the plant's past productivity. η_t indexes how the cost of building productivity varies across plants. The final identification result is the same regardless of whether productivity is chosen in period t (immediately productive) or in period $t - 1$ (time to develop).
- $G_t(K_{t+1}, K_t, \kappa_t)$ gives the cost of building a certain level of capital or capacity given the current level in period t . The dependence on past capacity allows for it to be more expensive to go from low to high capacity than from medium to high capacity. κ_t indexes how the cost of building capacity varies across plants.
- For simplicity, I assume plants believe they have perfect foresight⁷.

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

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

$$\max_{K,A,Q} \sum_{t=1}^T \beta_t \times \left[\underbrace{P_t(Q_t, \xi_t) Q_t}_{\text{Revenue}} - \underbrace{C_t\left(\frac{Q_t}{A_t}, K_t, W_t\right)}_{\text{Production costs}} - \underbrace{M_t(A_t, A_{t-1}, \eta_t)}_{\text{Technology cost}} - \underbrace{G_t(K_{t+1}, K_t, \kappa_t)}_{\text{Capital adjustment costs}} \right]. \quad (5)$$

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

$$0 \leq Q_t \leq K_t, \quad A_t \geq 0. \quad (6)$$

The model nests the basic proxy model developed in [Olley and Pakes \(1996\)](#). In the context of this model, the [Olley and Pakes \(1996\)](#) proxy model assumes productivity is *very* expensive to adjust (exogenous productivity)⁸, ξ_t does not vary across plants⁹, and there are no capacity constraints¹⁰.

2.1.1 Giving meaning to the parameters (ξ, η, κ)

I now put structure on the parameters of the plant's decision problem. My first task is to give the parameters ξ , η , and κ a scalar meaning as opposed to indexing an entirely general function space. Only then can we compare the demand curves, say, of high ξ plants to the demand curves for low ξ plants. Assumption 1 specifies the meaning of the parameters. In short, greater ξ means greater marginal revenue for a given output level, greater η means a lower marginal cost of productivity, and greater κ means a lower marginal cost of capital.

Assumption 1. Marginal revenue is increasing in ξ_t ,

$$\frac{\partial}{\partial \xi_t} \left[\frac{\partial P_t}{\partial Q} (Q_t, \xi_t) Q_t + P_t(Q_t, \xi_t) \right] \geq 0. \quad (7)$$

The marginal cost of productivity is decreasing in η_t ,

$$\frac{\partial M_t}{\partial A_t \partial \eta_t} (A_t, A_{t-1}, \eta_t) \leq 0, \quad (8)$$

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

⁹There is only one unobserved state variable allowed in the proxy model: productivity.

¹⁰Capacity constraints would violate the proxy assumption that input demand is increasing in productivity for fixed capacity. To see this, consider the [Levinsohn and Petrin \(2003\)](#) proxy: flexible inputs. The proxy model argues that $z_t = z_t(a_t, k_t)$ is increasing in a_t . For fixed capacity, increasing productivity eventually causes capacity constraints to bind. When capacity constraints bind, further increasing productivity will cause the plant to cut back on its flexible input use to keep its output below its capacity.

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

$$\frac{\partial M_t}{\partial A_{t-1} \partial \eta_t} (A_t, A_{t-1}, \eta_t) \leq 0. \quad (9)$$

The marginal cost of capital is decreasing in κ_t and the reduction in marginal cost from higher capital is also greater the greater κ_t (symmetric to the cost of productivity),

$$\frac{\partial G_t}{\partial K_{t+1} \partial \kappa_t} (K_{t+1}, K_t, \kappa_t) \leq 0 \quad (10)$$

$$\frac{\partial G_t}{\partial K_t \partial \kappa_t} (K_{t+1}, K_t, \kappa_t) \leq 0 \quad (11)$$

2.1.2 Convex marginal cost of the variable inputs

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

First, I assume the cost of the Z inputs (C) is (weakly) convex in the required output. This assumption does not rule out increasing returns to scale. For example, in a Cobb Douglas production function where Z is a single input, $Q = Z^{\theta_z} K^{\theta_k} A$. The cost function, conditional on K and A choices, is convex if $\theta_z \leq 1$ even if $\theta_z + \theta_k > 1$ ¹¹. I rule out that $\theta_z > 1$, or, more generally, I assume the returns to scaling Z are not increasing for fixed (K, A) . This captures the economic intuition of (weakly) diminishing marginal product in the variable inputs.

Second, I assume inputs are *normal*. Inputs are normal if their conditional input demand functions, $Z\left(\frac{Q}{A}, K, W\right)$, are increasing in output. By the envelope theorem, this assumption can be written as a restriction on the cost function,

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

The best argument for normal inputs is the behavior of non-normal inputs. Plants produce *more* output when the price of a non-normal input *increases*. It seems unlikely that an increase in fuel prices will cause power plants to produce more electricity, but that is what would happen if fuel were not a normal input. I assume inputs are normal using the principle that increases in a plant's cost of production will, all else equal, decrease how much a plant produces.

Lastly, I assume that, to the extent K is capital not capacity, it increases the marginal product of the variable inputs. The reasonableness of this assumption will vary from application-to-application, but capital is likely complementary with variable inputs in many production

¹¹Because productivity is also a factor of production, the returns to scale in this model also depend on the cost of productivity. It is not simply the sum of the output elasticities of the production function F .

problems. Better machines make people and materials more productive inputs. The Cobb Douglas production function always satisfies this restriction. If K is just capacity, then it will not affect the marginal product of the variable inputs (Z) at output levels below the capacity constraint, satisfying the restriction.

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

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

$$\frac{\partial^2 C_t}{\partial Q_t^2} \geq 0. \quad (13)$$

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

$$\frac{\partial Z_t \left(\frac{Q_t}{A_t}, K_t, W_t \right)}{\partial Q_t} \geq 0. \quad (14)$$

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

$$\frac{\partial F(Z_t, K_t)}{\partial K_t \partial Z_{t,\ell}} \geq 0 \quad \forall \ell = 1, \dots, L. \quad (15)$$

Lemma 1 establishes that Assumption 2 implies $\frac{\partial C}{\partial Q \partial K} C \left(\frac{Q}{A}, K, W \right) \leq 0$ which will be useful in developing the comparative static result when K is capital.

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

$$\frac{\partial C_t}{\partial Q_t \partial K_t} \left(\frac{Q_t}{A_t}, K_t, W_t \right) \leq 0. \quad (16)$$

Proof. See Appendix A. □

2.1.3 Convex investment costs for capital and productivity

G gives the cost of accumulating future capital (or capacity) as a function of past capital. It includes the cost of investing in new capital and the cost of maintaining present capital. One specification for the cost of investment part of G is,

$$\text{constant}_t \times (K_{t+1} - \delta \times K_t)^{\text{exponent}_t}, \quad (17)$$

where δ is the depreciation rate of capital.

I assume capital faces (weakly) convex adjustment costs. In the above example, that amounts to requiring $\text{exponent}_t \geq 1$. To make larger capital investments in a fixed amount of time, the plant must spend more for each additional unit of investment. It is progressively more expensive to purchase larger or better machines and more expensive to finance large expenditures than small ones. Intuitively, the interest rate on large loans is greater than the interest rate on small loans so the cost of raising twice as much capital is more than double.

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

I make the following mathematical assumption to capture the economic assumption of convex investment costs.

Assumption 3. The cross partial of G_t is negative for all κ_t ,

$$\frac{\partial G_t}{\partial K_t \partial K_{t+1}}(K_{t+1}, K_t) \leq 0 \quad (18)$$

The cross partial of M_t is negative for all η_t ,

$$\frac{\partial M_t}{\partial A_t \partial A_{t-1}} \leq 0. \quad (19)$$

We can see that this assumption holds for, say, $G_t = \text{constant}_t \times (K_{t+1} - K_t)^{\text{exponent}_t} - \kappa_t K_t$ (the second term is the cost of maintaining current capital), when investment costs are convex,

$$\frac{\partial G_t}{\partial K_t \partial K_{t+1}} = -\text{constant}_t \times (\text{exponent}_t - 1) \times \text{exponent}_t \quad (20)$$

$$\times (K_{t+1} - K_t)^{\text{exponent}_t - 2} \leq 0. \quad (21)$$

2.1.4 The revenue maximization problem has no interior solution

I have put structure on the cost side of the plant's problem (C , G , and M). I now add structure to the demand side. I assume the revenue maximization problem has no interior solution. There is no finite revenue-maximizing level of output ignoring the costs of producing it¹². Costs limit plant output not market power. While this may not be the case for some products, like pharmaceuticals, it seems likely to be the case for many other products which either have sufficient competition so that the residual demand curve is not too inelastic or

¹²This assumption is also necessary in the proxy model. If there were a revenue-maximizing output, then input demand could not be strictly increasing in productivity as required in the proxy model. For some level of productivity, output would reach the revenue-maximizing level and beyond that point, greater productivity would cause the plant to cut back on other inputs to maintain output at the revenue-maximizing level.

that demand for the product itself is elastic enough. We can operationalize this economic assumption with the following restriction on the plant's (residual) demand curve:

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

$$\frac{\partial}{\partial Q} [\text{MR}(Q, \xi) Q] \geq 0 \quad \text{for all } \xi. \quad (22)$$

This assumption is true in a wide variety of demand models. It is true when the plant's demand curve has a constant elasticity, when demand is logistic (as in standard discrete choice models), when plants take prices as given, and for any elastic demand curve where the absolute value of the elasticity of demand is increasing in price¹³.

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

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

Proof. See Appendix A. □

2.1.5 Main comparative static result

Given this structure, I derive a comparative static result. The choice of productivity, capital or capacity, and output is increasing in the underlying plant-specific state variables (Theorem 2). The form of the result differs slightly depending on whether K is capital or whether it is capacity.

¹³To see this,

$$\begin{aligned} \text{MR}(Q, \xi) Q &= \underbrace{P(Q, \xi) Q}_{\text{Revenue}} \times \underbrace{\left[\frac{P'(Q, \xi)}{P(Q, \xi)} \times Q + 1 \right]}_{\text{Inverse demand elasticity plus one}} \\ \Rightarrow \frac{\partial}{\partial Q} [\text{MR}(Q, \xi) Q] &= \text{MR}(Q, \xi) \times \frac{\partial}{\partial Q} \left[\frac{P'(Q, \xi)}{P(Q, \xi)} \times Q \right] \\ &\quad + \text{Revenue}(Q, \xi) \times \left[\frac{P'(Q, \xi)}{P(Q, \xi)} \times Q + 1 \right] \end{aligned}$$

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

Theorem 2. Let $Y^T = (Y_1, \dots, Y_T)$ for any variable Y . Define $s = (\eta^T, \xi^T, \kappa^T, K_0, A_0)$ as the vector of plant-specific state variables. Given Assumptions 1, 2, 3, and 4, then

- (1) If K is capital, so that $Q = F(Z, K)A$,
 - (2) OR if K is capacity, so that $Q = \min\{F(Z)A, K\}$,
- then $A_t(s, -W^T)$, $Q_t(s, -W^T)$, and $K_t(s, -W^T)$ are increasing functions of s for fixed W^T , and, if K is capacity, then $A_t(\cdot)$, $Q_t(\cdot)$, and $K_t(\cdot)$ are increasing in all arguments, including W^T

Proof. See Appendix A. □

Theorem 2 establishes that when the state variables increase, plants choose greater output, productivity, and capacity. When the plant is incentivized to produce more output, it is also incentivized to be more productive and to choose greater capacity.

2.2 Markovian state variables and positive association

$\tilde{s} = (\xi^T, \eta^T, \kappa^T, -W^T, K_0, A_0)$ is the vector of unobserved state variables¹⁴ that (potentially) vary across plants. If there was only one plant-specific unobserved state variable (if \tilde{s} was a scalar), then from the above comparative static result, $A_t(\tilde{s})$, $K_t(\tilde{s})$, and $Q_t(\tilde{s})$ would be positively correlated because increasing functions of the same random variable are positively correlated. But \tilde{s} is not a scalar. I show that productivity, capacity, and output are positively related for multi-dimensional \tilde{s} if we add a mild statistical structure on how \tilde{s} varies across plants.

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

Assumption 5. Define $\tilde{s}_t = (K_0, A_0, \xi_t, \eta_t, \kappa_t, -W_t)$. Let j index the individual elements of \tilde{s}_t . State variables evolve according to the rule,

$$\tilde{s}_{j,t} = \sigma_{j,t}(\tilde{s}_{t-1}, v_{j,t}), \quad (23)$$

where $\sigma_{j,t}$ is a weakly increasing function for all (j, t) .

Define $\tilde{s}^T = \cup_{t=0}^T \tilde{s}_t$. \tilde{s}^{t-1} and $v_t = (v_{1,t}, \dots, v_{J,t})$ are independent.

¹⁴The fact that this approach can deal with unobserved input prices (W) is a nice property. The proxy model is not identified with unobserved input price variation, see Flynn, Gandhi, and Traina (2019).

Definition 1: Positive Association

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

Let X be a random vector. If for any two increasing (in all arguments) functions, ϕ_1 and ϕ_2 , such that the required expectations exist,

$$\text{cov}[\phi_1(X), \phi_2(X)] \geq 0, \quad (24)$$

then X is a positively associated random vector.

Positively associated random vectors have the following properties:

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

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

The assumption that $\sigma_{j,t}$ is increasing nests the standard Markov structure, $\tilde{s}_{j,t} = \sigma_{j,t}(\tilde{s}_{j,t-1}, v_{j,t})$, where greater $\tilde{s}_{j,t-1}$ implies a larger distribution of $\tilde{s}_{j,t}$ (persistence) but adds a bit of generality by allowing the other lagged state variables to positively affect another state variable. The assumption that the (nonseparable) shocks are independent of the prior state is the Markovian assumption. For example, if the state variables follow an AR1 process (with time-varying drift and persistence parameters), they satisfy Assumption 5.

Next, I restrict how the shocks $(v_{1,t}, \dots, v_{J,t})$ relate to one another. I assume they are *positively associated*. A random vector is positively associated if the covariance of any two increasing functions of the vector is positive. It is a stronger restriction than positive correlation but a weaker restriction than statistical independence. Definition 1 states some useful properties of positive association.

Positive association is credible. It encapsulates standard modeling assumptions: if the shocks $(v_{1,t}, \dots, v_{J,t})$ are independent of each other, they are positively associated. But positive association also allows general *ability* shocks to introduce positive dependence between the shocks.

Consider the following example model of v ,

$$v_{j,t} = \iota_j \epsilon_t + e_{j,t} \quad (25)$$

where $e_{j,t}$ is independent of $e_{k,t}$ and $\iota_j > 0$ is a fixed parameter. All of the state variables are “good” for the plant in the sense that they make the plant more profitable. So this structure allows for a “talent” or quality of the plant shock in addition to shocks that affect the specific variables. The covariance of the shocks is then, $\text{cov}(v_{j,t}, v_{k,t}) = \iota_j \iota_k \text{var}(\epsilon_t) \geq 0$. I assume $(v_{1,t}, \dots, v_{J,t})$ is positively associated to capture correlations introduced in this way

instead of just making the stronger (but still reasonable) assumption that the shocks are independent.

Note that this structure puts no restriction on the dynamics of v_t . The shocks do not need to be independent or identically distributed across time. It is only a restriction on the way the shocks relate to one another in a given time period.

Assumption 6. $v_t = (v_{1,t}, \dots, v_{J,t})$ is a positively associated vector.

I use assumptions 5 and 6 and apply the properties of positive association to show that once the state variables are positively associated, they will always be (Lemma 2).

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

Proof. See Appendix A. □

Lastly, I assume that we have already reached this absorbing state in the data. If the initial state variable draws are, say, independent of each other, a natural assumption, then we would always be in the absorbing state. I make the weaker assumption that at some point prior to the data the state variables were positively associated. This rules out only that the initial state variables were negatively correlated with each other and that negative relationship has not diminished over time.

Assumption 7. $(\xi_0, \eta_0, \kappa_0, A_0, K_0)$ is positively associated.

Theorem 3 concludes \tilde{s} itself is positively associated by induction.

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

Proof. See Appendix A. □

2.2.1 The Markov structure when K is capital

When K is capital, we need the additional assumption that $-W^T$ is independent of s because we need that s is positively associated conditional on W^T for the identification result to hold. We can think of this requirement as imposing further structure on the Markov process for W_t . Essentially, I assume that s_t (the non-input price state variables) and W_t evolve following separate Markov processes.

In particular, I assume W_t 's Markov process depends only on its own lagged values,

$$-W_t = \sigma_{W,t}(-W_{t-1}, v_{W,t}), \quad (26)$$

and that s_t (the non-input price state variables) do not depend on lagged input prices,

$$s_t = \sigma_{s,t}(s_{t-1}, v_{s,t}), \quad (27)$$

and, lastly, to make the processes separate,

$$v_{W,t} \text{ is independent of } v_{s,t} \quad (28)$$

$$s_0 \text{ is independent of } -W_0. \quad (29)$$

This additional structure implies s *conditional on* W^T is positively associated. To see this, note the following two points:

- (1) Because Theorem 3 has already established that $(s, -W^T)$ is positively associated, it follows that s is positively associated (unconditional on W^T) because any subset of a positively associated random vector is positively associated.
- (2) The above structure implies that W^T is independent of s .

Therefore, for any two increasing functions ϕ_1 and ϕ_2 ,

$$\text{cov}[\phi_1(s), \phi_2(s) | W] = \text{cov}[\phi_1(s), \phi_2(s)] \geq 0, \quad (30)$$

where the equality holds by the independence of W and s and the inequality by the result of Theorem 3.

2.3 Identified set

Positive association is an ordinal property¹⁵: increasing functions of random vectors that are positively associated are themselves positively associated. Because the plant's choices $A_t(s, -W^T)$, $K_t(s, -W^T)$, and $Q_t(s, -W^T)$ are increasing in $\tilde{s} = (s, -W^T)$ and \tilde{s} is positively associated, (A_t, K_t, Q_t) are positively associated.

With data on (Q_t, K_t, Z_t) some production functions imply distributions of A_t that satisfy this restriction and others do not. Requiring the production function to satisfy this restriction partially identifies the production function (Theorem 4).

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

$$\begin{aligned} \mathcal{F} = \{f \in \text{increasing functions} : & \text{for all functions } \phi_1 \text{ and } \phi_2 \\ & \text{that are increasing in all arguments,} \\ & \text{cov}[\phi_1(q - f(z, k), q, k), \phi_2(q - f(z, k), q, k)] \geq 0\}. \end{aligned} \quad (31)$$

¹⁵Unlike other forms of positive dependence between random variables. For example, Manski and Pepper (2000) propose monotone instrumental variable assumptions which suppose the following form of positive dependence between covariates and the residual: $\partial \mathbb{E}[x_1 | x_2] / \partial x_2 \geq 0$. 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 $x_1 \sim N(x_2, 2x_2^{-1})$ 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.

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

Proof. By Theorem 3, \tilde{s} is positively associated. By Theorem 2, $a(\tilde{s})$, $q(\tilde{s})$, and $k(\tilde{s})$ are all increasing functions. By Property 2, increasing functions of positively associated random variables are positively associated. Therefore, (a, q, k) are positively associated. The proof when K is capital is identical, except conditional on W^T . \square

3 Inference on productivity regressions

I develop a practical method to use data on (Q, K, Z) and the restriction that $f \in \mathcal{F}$ to form bounds on a common class of parameters of interest in the productivity literature. I make inference on coefficients in regressions of productivity on plant-level characteristics and policy variables instead of directly on the production function itself. We are not usually interested in the productivity of any particular plant but in how productivity varies across plants of different types or that were exposed to different policies. Focusing directly on the final parameter of interest simplifies inference.

I consider estimating bounds on the parameter τ in the linear (instrumental variable) regression where log productivity is the dependent variable,

$$a_t = x_t^\top \beta + u_t \tag{32}$$

$$\mathbb{E}[v_t u_t] = 0 \tag{33}$$

$$\mathbb{E}[v_t x_t^\top] \text{ is left-invertible} \tag{34}$$

$$\tau = \rho^\top \beta, \tag{35}$$

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

Problems of this form are common in the productivity literature. For some examples of what τ can be, Pavcnik (2002) estimates the effect of trade liberalization in Chile on productivity using a specification with this form, Olley and Pakes (1996) answer how productivity in the telecommunications equipment industry changed after the industry restructured, and De Loecker (2011) studies how trade liberalization affected productivity in the Belgian textile

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

industry¹⁷. τ is the effect of interest in all these articles. In this article, τ is the effect of restructuring on the productivity of power plants in the electricity generation industry.

Within the model of Section 2, x_t and v_t can be viewed as observable determinants or correlates of the plant's fundamental parameters (ξ, η, κ) . Our goal is to understand how, say, restructuring affected these parameters to increase or decrease plant productivity.

I focus directly on bounding the final parameter of interest. The advantage of doing so is best seen by considering the alternative. The alternative is to form a confidence set for the production function and search over the production functions in that set, computing τ for each, a computationally intensive task. By focusing directly on the final parameter of interest, we can simplify computation of the identified set for τ . This strategy is common in the econometrics literature for this reason, see [Kaido, Molinari, and Stoye \(2016\)](#) and [Bugni, Canay, and Shi \(2016\)](#).

We can write τ as,

$$\tau = \rho^\top \mathbb{E} \left[v_t x_t^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t a_t] \quad (36)$$

$$= \rho^\top \mathbb{E} \left[v_t x_t^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t q_t] - \rho^\top \mathbb{E} \left[v x^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t f(z_t)]. \quad (37)$$

The first term is point-identified from the data. In fact, it is the estimate of “ τ ” when we treat q_t as the dependent variable in the regression instead of a_t , as if we were estimating the following regression,

$$q_t = x_t^\top \beta(q) + u_t(q), \quad \mathbb{E} [v_t u_t(q)] = 0 \quad (38)$$

$$\tau(q) = \rho^\top \beta(q). \quad (39)$$

More generally, let $\tau(y)$ be the estimate of τ when the above regression model is used with the dependent variable y . We can write the parameter of interest as $\tau = \tau(q) - \tau(f)$. τ is partially identified because f is partially identified so $\tau(f)$ is only partially known. Bounds on τ are then, by definition,

$$\left[\min_{f \in \mathcal{F}} \tau(q) - \tau(f), \max_{f \in \mathcal{F}} \tau(q) - \tau(f) \right]. \quad (40)$$

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

This section has two parts. In the first, I derive a simple-to-compute form for the identified set. In the second, I propose a method to make inference on the identified set.

3.1 Linear positive association: a convex identified set

The identified set \mathcal{F} is not convex so the bounds on τ are the values of two nonconvex programming problems. Such problems are difficult to solve. I derive a relaxation of the

¹⁷For a larger list of problems like these, see [Syverson \(2011\)](#)'s literature review.

full positive association assumption that is convex, making the bounds easier to compute. There are well-established algorithms for solving convex programming problems that are guaranteed to converge in polynomial time¹⁸.

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

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

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

I introduce a convex relaxation of positive association I call “linear positive association”. It uses a subset of the ϕ functions used to construct the positive association identified set, \mathcal{F} . This results in a larger identified set, but the set is easier to compute. Linear positive association identifies the log production function in the following set, where Df denotes the derivative of f in each of its arguments,

$$\mathcal{F}_{\text{LPA}} = \left\{ f : \underbrace{Df \geq 0}_{\text{Increasing production function}}, q - f = a, \underbrace{a \geq 0}_{\text{Constant normalization}}, \right. \\ \left. \underbrace{\text{cov}[a\phi_1(q, k), \phi_2(q, k)] \geq 0}_{\text{Linear positive association}} \right. \quad (41) \\ \left. \text{where } \phi_1 \geq 0 \text{ and } \phi_2 \text{ are increasing.} \right\}.$$

Linear positive association is weaker than positive association (given Assumption 8) because an increasing function f belongs to \mathcal{F}_{LPA} if it implies positive covariance between a certain class of increasing functions of (a, k, q) instead of all increasing functions of (a, k, q) as required to belong to \mathcal{F} . But \mathcal{F}_{LPA} is convex, an attractive property for computing the bounds.

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

Proof. See Appendix A. □

Bounds on τ given that $f \in \mathcal{F}_{\text{LPA}}$ can be computed by solving two convex programming problems, which simplifies the computational problem,

$$\max \text{ or } \min_{f \in \mathcal{F}_{\text{LPA}}} \mathbb{E} \left[vx^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t q_t] - \mathbb{E} \left[vx^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t f(z_t)]. \quad (42)$$

¹⁸For example, the log barrier method.

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

In practice, we usually assume a linear-in-parameters functional form for the production function, like the Cobb Douglas or translog production function, justifying them as first or second order approximations. Linear positive association is especially attractive in this case because the bounds can be written as the values of two linear programming problems.

Suppose the production function is,

$$f(z) = r(z)^\top \theta, \quad (43)$$

where $r(z)$ is a known vector of functions and θ is an unknown parameter vector. Linear positive association requires θ to satisfy infinitely many linear inequalities,

$$\Theta_{\text{LPA}} = \left\{ \theta : \underbrace{Dr\theta \geq 0}_{\text{Increasing production function}}, q - r(z)^\top \theta = a, \underbrace{a \geq 0}_{\text{Constant normalization}}, \right. \\ \left. \underbrace{\text{cov}[a\phi_1(q, k), \phi_2(q, k)] \geq 0 \quad \forall (\phi_1, \phi_2) \text{ where } \phi_1 \geq 0, D\phi_1 \geq 0, D\phi_2 \geq 0}_{\text{Linear positive association}} \right\} \quad (44)$$

While we could use all of these constraints in estimation, some of the constraints would then have substantial identifying power via the functional form we assume for f . I propose using only a subset of the constraints based on the functional form of the production function.

Empirical economics commonly follows this principle when estimating regressions. Consider a regression problem of the form $y = \mathbb{E}[y|x] + \epsilon$. If we approximate $\mathbb{E}[y|x]$ by $\beta_0 + \beta_1 x$ we only use the moment restrictions $\mathbb{E}[\epsilon] = \mathbb{E}[x\epsilon] = 0$. But if we approximate $\mathbb{E}[y|x]$ by $\beta_0 + \beta_1 x + \beta_2 x^2$ we use the moments $\mathbb{E}[\epsilon] = \mathbb{E}[x\epsilon] = \mathbb{E}[x^2\epsilon] = 0$. Even though in both cases x^2 could be used as an instrument if we assume $\mathbb{E}[\epsilon|x] = 0$. We do so because we recognize our functional form as an approximation. Similarly, in this setting, with a less flexible production function, highly flexible ϕ functions may unduly restrict the identified set simply because it is difficult to fit a line, say, underneath a very concave curve.

To be conservative and avoid obtaining narrow bounds through functional form, I only use increasing functions ϕ_1 and ϕ_2 that are similar in complexity to the production function itself. For example and because this is the functional form I use in my application, if the production function is Cobb Douglas,

$$r(z) = (1, z), \quad (45)$$

I use only ϕ_1 and ϕ_2 that are also linear in logs.

$$\Theta_{\text{CD,LPA}} = \{ \theta : \theta_z \geq 0, q - \theta_1 - z^\top \theta_z = a, a \geq 0, \\ \text{cov}[a\phi_1(q, k), \phi_2(q, k)] \geq 0, \\ \phi_1, \phi_2 \in \{ \phi : \phi = \varphi_1 + \varphi_k k + \varphi_q q, \varphi_k \geq 0, \varphi_q \geq 0, \varphi_1 + \varphi_k k + \varphi_q q \geq 0 \} \} \quad (46)$$

Theorem 6 shows the set $\Theta_{\text{CD,LPA}}$ is in fact the intersection of only a finite number of linear inequalities.

Theorem 6. Assume that the support of the observed distribution of (q, z, k) is a subset of the rectangle: $[q, \bar{q}] \times [z_1, \bar{z}_1] \times \cdots \times [z_L, \bar{z}_L] \times [\underline{k}, \bar{k}]$. $\Theta_{\text{CD,LPA}}$ can be equivalently expressed as,

$$\begin{aligned} \Theta_{\text{CD,LPA}} = \{ \theta : & \underbrace{\theta_z \geq 0}_{\text{Increasing production function}}, q - \theta_1 - z^\top \theta_z = a, \\ & \underbrace{q - \theta_1 - \bar{z}^\top \theta_z \geq 0}_{\text{Constant normalization}}, \\ & \underbrace{\text{cov}[a\phi_1(q, k), \phi_2(q, k)] \geq 0}_{\text{Linear positive association}}, \\ & \phi_1, \phi_2 \in \left\{ \phi : \phi = (k - \underline{k}), \phi = (q - \underline{q}), \text{ or } \phi = 1 \right\} \} \end{aligned} \quad (47)$$

Proof. See Appendix A. □

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

$$\max \text{ or } \min_{\theta \in \Theta_{\text{CD,LPA}}} \mathbb{E} \left[vx^\top \right]_{\text{left}}^{-1} \mathbb{E} [v_t q_t] - \mathbb{E} \left[vx^\top \right]_{\text{left}}^{-1} \mathbb{E} \left[v_t \left(1, z_t^\top \right)^\top \right] \theta. \quad (48)$$

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

But we do not know the parameters of these linear programming problems. We have to estimate them from data and account for estimation error.

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

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

Because the vector μ is a reduced-form parameter, Bernstein-von Mises theorems imply the posterior distribution of μ converges to a normal distribution centered around the true

parameter value, see [Bickel and Kleijn \(2012\)](#) and the results they cite. The convergence is in total variation,

$$\sup_h \left| \int h d\pi(\mu|\text{data}) - \int h d\Phi_n(\mu) \right| \rightarrow 0, \quad (49)$$

regardless of the prior specified for μ ¹⁹. This result allows us to use the distribution $\Phi_n(\mu)$ as an approximation to the posterior distribution $\pi(\mu|\text{data})$ in large samples without specifying a prior for μ directly.

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

$$\rho_n(\tau_0) = \int_{\mu} \mathbf{1}\{\tau_0 \in [\tau_{LB}(\mu), \tau_{UB}(\mu)]\} d\Phi_n(\mu), \quad (50)$$

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

I use the $\rho_n(\cdot)$ function to form what I call a “Bayesian Hypothesis Interval” (BHI). We can construct frequentist confidence intervals by searching for parameter values that are not rejected by the α hypothesis test. The BHI finds 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 another common inference method, Bayesian credible intervals. Credible intervals are any interval such that our posterior belief the parameter belongs to it is greater than a certain value. So while hypothesis intervals are uniquely defined, there are many valid credible intervals of a given size. More critically, credible intervals that cover τ depend on the prior specified for τ even in very large samples. τ is only partially identified so the posterior distribution does not concentrate around a single point. Hypothesis intervals only depend on the posterior distribution for the point-identified parameters τ_{LB} and τ_{UB} .

The hypothesis interval is defined as,

$$HI_{\alpha,n} = \{\tau_0 : \rho_n(\tau_0) \geq \alpha\}. \quad (51)$$

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

¹⁹So long as the prior’s support includes the reals or otherwise includes a neighborhood of the (frequentist) probability limit of $\hat{\mu}$

To compute the hypothesis interval, I do the following:

- (1) Take B draws from the distribution $N(\hat{\mu}, n^{-1}\hat{\Sigma})$: $\mu^{(1)}, \dots, \mu^{(B)}$.
- (2) For each $\mu^{(b)}$, compute: $\tau_{LB}^{(b)} = \tau_{LB}(\mu^{(b)})$ and $\tau_{UB}^{(b)} = \tau_{UB}(\mu^{(b)})$ by solving the two linear programs.
- (3) The hypothesis interval includes all τ_0 such that:

$$\frac{1}{B} \sum_{b=1}^B \mathbf{1}(\tau_{LB}^{(b)} \leq \tau_0 \leq \tau_{UB}^{(b)}) \geq \alpha \quad (52)$$

3.4 Why I use a Bayesian method

Bayesian inference has practical advantages over frequentist inference in partially identified models. When the bounds on the parameter of interest are non-differentiable but known functions of point-identified parameters, Bayesian methods can simply sample from the posterior distribution for the point-identified parameters and apply the functions to recover posterior beliefs about the parameter of interest. $\tau_{LB}(\mu)$ and $\tau_{UB}(\mu)$ are examples of such functions. The non-smoothness of $\tau_{LB}(\mu)$ and $\tau_{UB}(\mu)$ poses no difficulty for Bayesian inference because the posterior is conditional on the data. We do not need to consider how the distribution of the estimated parameters μ changes as sample size increases. But the asymptotic distribution for frequentist inference depends on exactly that, complicating frequentist inference on non-smooth functions of means²⁰.

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

4 Why I do not use a proxy or instrumental variable approach

In this short section, I discuss why I introduce a new method for estimating endogenous productivity. I explain why I do not pursue a modification of the proxy approach or search for an instrument.

²⁰[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.

The main method in the current literature for estimating endogenous productivity models is from [Doraszelki and Jaumandreu \(2013\)](#) and [De Loecker \(2013\)](#) who use a modification of the proxy model. The following transition equation in their model illustrates their main identification assumptions,

$$a_t = g(a_{t-1}, x_{t-1}) + e_t, \quad \mathbb{E}[e_t | a_{t-1}, x_{t-1}, k_t] = 0 \quad (53)$$

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

This structure rules out much of what the general model of productivity choice I propose allows. In many interesting models of productivity, e_t as defined above will either be partially known in $(t-1)$ or x_{t-1} will be chosen partially in period t . For example, the following are ruled out by the proxy approach, but allowed in the comparative static approach²¹:

- **Flexible Productivity.** Productivity is flexible if it can be adjusted in the current period. This is ruled out by the proxy model because either (1) x_{t-1} includes this flexible choice, in which case it is correlated with e_t because x_{t-1} then must be in the period t information set or (2) x_{t-1} does not control for this choice so the choice is a part of e_t which is then chosen with knowledge of the plant's current state variables, among which are (a_{t-1}, x_{t-1}, k_t) violating the moment condition. An example of flexible productivity is workers putting in greater effort because a firm introduced a new performance incentive. Flexible productivity choice is an especially acute example of the proxy model's identification problems with flexible factors of production. For the identification problem with *observed* flexible inputs, see [Gandhi, Navarro, and Rivers \(2019\)](#).
- **Unobserved Adjustments to Productivity.** Even if productivity is inflexible and a firm makes all its decisions about productivity in period $t-1$, if some of that investment is not included in x_{t-1} , the proxy model will not be identified. If x_{t-1} does not completely control for prior investment in productivity, some of that investment is included in e_t and that investment was made with knowledge of the $(t-1)$ information set. e_t will then be correlated with variables in the $(t-1)$ information set, like (a_{t-1}, x_{t-1}, k_t) , contrary to the assumed moment conditions.

This is important because it may seem like when we want to understand the causal effect of something on productivity we can just include the variable of interest in x_{t-1} , but we cannot. For example, [De Loecker \(2013\)](#) sets x_{t-1} to whether the plant is an exporter in period t . He wants to find the causal effect of exporting on productivity, not

²¹Here I focus on the differences in the model of productivity not on the other generalizations I make relative to the proxy model. For example, I also allow for multiple unobserved state variables that cannot be proxied by observables and capacity constraints which are ruled out by the proxy model.

just the selection effect described in Melitz (2003). It is very reasonable to assume that entry into exporter status in period t is uncorrelated with information that only became available in period t . But if we think plants can learn from exporting, we should also think plants can learn in other ways or, more generally, that plants can take actions that affect their productivity. Unless exporter status *fully* explains plant investment decisions in productivity, the proxy approach does not identify the production function.

Direct instrumental variable approaches also face additional challenges when productivity is a factor of production. I put to the side the important data issues with finding instruments in common production datasets and consider the general instrumental variable problem. The production equation is, using the notation from the model in Section 2,

$$q_t = f(z_t(\tilde{s}), k_t(\tilde{s})) + a_t(\tilde{s}). \quad (54)$$

(z, k) are determined by the exact same state variables as a . The only difference between productivity and inputs is that we observe inputs and do not observe productivity. Therefore, finding a variable that is correlated with one set of factors of production (observed inputs) but not correlated with the choice of another factor of production (productivity) requires imposing restrictions that make productivity in some way different than the other factors of production. For example, a demand shock (a change in ξ) would naturally affect the choice of *all* factors of production, observed inputs and productivity. A shock to the price of an observed input would cause substitution to or from productivity like it would for all other factors of production.

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

5 How restructuring in the electricity industry affected productivity

I now show that the bounds developed in the previous sections are narrow enough for us to learn about real empirical problems. I study how restructuring in the electricity industry affected power plant productivity using the approach developed in this article. This application also demonstrates the empirical importance of allowing for endogenous productivity because the main mechanism through which restructuring might affect a plant's productivity is by affecting its incentive to be more productive.

5.1 Industry and policy background

I start by briefly describing the history of regulation in the electricity industry, how the restructuring policy movement changed the electricity generation industry in the United States, and how this policy change affected the incentives of power plants to be more productive.

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 investments²². 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 markets determine the price of electricity. By disintegrating the utilities, they were able to regulate the naturally monopolistic stage of electricity production, electricity transmission, while allowing for competition in the generation stage. I study how this policy affected power plant productivity. See [Borenstein and Bushnell \(2015\)](#) for a fuller history of the restructuring policy.

Market-based pricing gives power plants an incentive to be more productive because, if they have lower costs than their competitors, they can earn greater markups. But the restructuring policy 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, *lower* productivity may be associated with an *increase* in welfare (see Section 5.9). 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 empirically. 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.

5.1.1 How power plants might choose their productivity

There is some 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

²²In 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.

this is possible suggests that there are actions operators can take that increase the efficiency of the power plant. Additionally, in interviews, they found that people in the industry believe the “skill and effort” of employees (a form of productivity) make 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 \(2009\)](#)

5.2 Data

To use the approach in this article, I need data on power plant output and input use. I use the same data as [Fabrizio, Rose, and Wolfram \(2007\)](#) for the years 1981 to 1999²³, and then extend the dataset until 2003. The first restructuring policy changes were in 1996 so the data has information on power plants before and after some states restructured their electricity markets.

The data is at the power plant-level and includes large (greater than 100 megawatt capacity), fossil fuel power plants. I measure output by millions of megawatt-hours of electricity generated net of electricity the power plant uses itself, fuel use by units of heat energy (in millions of million British thermal units), capacity by the total nameplate capacity of the plant (in megawatts), labor by total employment, and non-fuel expenditures in millions of dollars. The data is annual.

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

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.

²³I thank Fabrizio, Rose, and Wolfram for making their data available and easy to use.

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

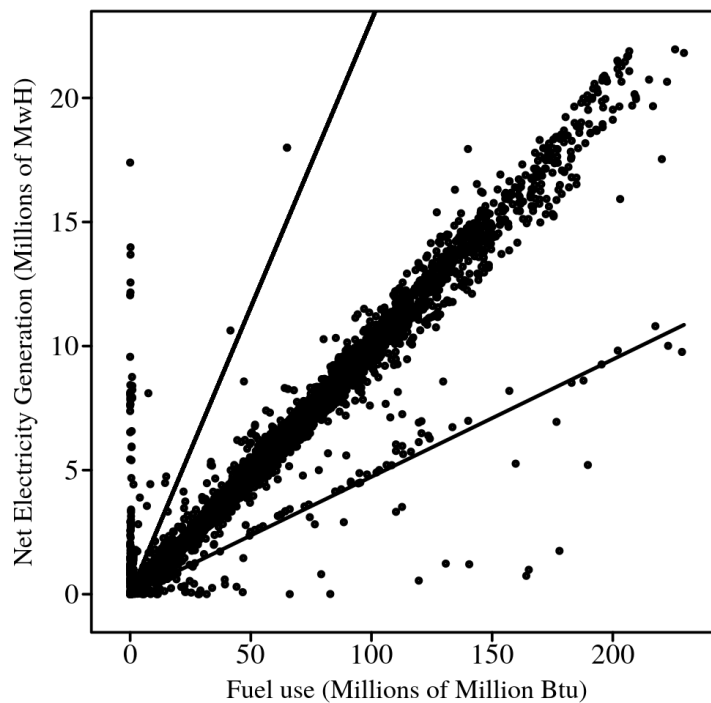


Table 1: Descriptive statistics of data (with mismeasured observations removed)

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.

5.3 A model of electricity production and a measure of productivity

I use the data to estimate the following model of electricity production.

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. 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 . The above description of electricity production translates into the following production equation (it is a version of the capacity model of K in Section 2),

$$Q^{(\sigma)} = \min \left\{ H^{(\sigma)} F(L, E) A, K^{(\sigma)} \right\}, \quad (55)$$

where L is the number of workers employed and E is nonfuel expenditures (mostly maintenance and other operating costs).

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,

$$\sum_{\sigma} Q^{(\sigma)} = F(L, E) \sum_{\sigma} H^{(\sigma)} A \implies Q = F(L, E) H A, \quad (56)$$

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

I call A “capacity-agnostic productivity” because it measures how output would vary across plants if all plants used the same inputs and were never capacity-constrained. For the purposes of constructing the identified set, we can treat (Q/H) (fuel efficiency) as our output measure and (L, E) as our vector of inputs (Z from the theory in Section 2),

$$\frac{Q}{H} = F(L, E) A. \quad (57)$$

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

I use a Cobb Douglas functional form for F ,

$$\frac{Q}{H} = L^{\theta_L} E^{\theta_E} A. \quad (58)$$

5.4 Identifying the effect of restructuring

With a measure of productivity in hand, I now turn to the problem of identifying how restructuring affected it. This identification problem is essentially orthogonal to the problem of identifying productivity itself. Even if we observed productivity in a dataset, we would need to deal with the problem that 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. Throughout this subsection, I treat productivity as known and consider the problem of identifying the effect of restructuring in isolation. I establish a mapping between a distribution of productivity and the effect of restructuring. In later sections, I will then use the partial identification result and the data to restrict the set of allowed productivity distributions and so, using this mapping, the effect of restructuring.

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

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

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

$$\begin{aligned} a_{it} = & \tau_a \times \text{Restruct}_{it} + \tau_{POST,a} Y_{it} \mathbf{1}(Y_{it} > 0) + \tau_{PRE,a} Y_{it} + \kappa_a m_{rt} \\ & + \mu_{20} \times \mathbf{1}(\text{Age}_{it} \leq 20) + \mu_{30} \times \mathbf{1}(\text{Age}_{it} \leq 30) \\ & + \mu_{40} \times \mathbf{1}(\text{Age}_{it} \leq 40) + \iota \times \text{EverRestruct}_{it} + \xi_i + \delta_t + u_{it}, \end{aligned} \quad (59)$$

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

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

The inclusion of the plant level fixed effects is important for interpreting the estimated effect. τ_a 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_a . A policy can have a selection effect without power plants having any control over their productivity. On the other hand, if we think power plants decide to be more or less productive in response to the new incentives offered by restructuring, then plants must be able to adjust their productivity. The plant-level effect of restructuring demonstrates the empirical relevance of allowing for productivity choice.

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

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

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.

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.

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

$$\begin{aligned}
a_{it} = & \tau_a \times \text{Restruct}_{it} \text{IO}_{it} + \beta_a \times \text{Restruct}_{it} + \tau_{POST,a} Y_{it} \mathbf{1}(Y_{it} > 0) \text{IO}_{it} + \tau_{PRE,a} Y_{it} \text{IO}_{it} \\
& + \beta_{POST,a} \times Y_{it} \mathbf{1}(Y_{it} > 0) + \beta_{PRE,a} \times Y_{it} + \kappa_a m_{rt} \\
& + \mu_{20} \times \mathbf{1}(\text{Age}_{it} \leq 20) + \mu_{30} \times \mathbf{1}(\text{Age}_{it} \leq 30) + \mu_{40} \times \mathbf{1}(\text{Age}_{it} \leq 40) \\
& + \iota_a \times \text{EverRestruct}_{it} + \iota_{IOU,a} \times \text{EverRestruct}_{it} \text{IO}_{it} + \xi_i + \delta_t + u_{it},
\end{aligned} \tag{60}$$

where IO indicates whether the plant is investor-owned. This specification measures the causal effect of restructuring by how productivity 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.

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

Having established a method of estimating the effect of restructuring on productivity for a given productivity distribution, I return to the problem of identifying productivity and so, the effect of restructuring on it. Recall from Section 3, that τ_a can be written as a weighted combination of the coefficients from a regression with output and inputs as the dependent variable with the same right hand side. From the model of electricity production proposed above,

$$\tau_a = \tau_{q-h} - \tau_\ell \theta_\ell - \tau_e \theta_e, \tag{61}$$

where τ_{q-h} is the effect of restructuring on fuel efficiency, τ_ℓ the effect on labor use, and τ_e the effect on nonfuel expenditures. The τ parameters are all point identified, but the θ are only partially identified.

Table 2 presents estimates of these parameters. 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 restructuring caused power plants to use less inputs and reduce the amount of electricity they produced (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.

With estimates for τ_{q-h} , τ_ℓ , and τ_e in hand, I use the partial identification method proposed in the previous sections of this article to restrict θ , the parameters of the production function. Doing so will partially identify τ_a .

5.6 How linear positive association identifies the effect of restructuring on productivity

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

$$\tau_a = \tau_{q-h} - \tau_f \geq \tau_{q-h}. \quad (62)$$

Because, from Section 5.5, $\tau_{q-h} < 0$, the empirical question is whether (θ_ℓ, θ_e) can be large enough that $(\tau_\ell \theta_\ell + \tau_e \theta_e)$ is greater in magnitude than τ_{q-h} or whether the linear positive association assumption sufficiently restricts how large θ can be so that $\tau_f = \theta_\ell \tau_\ell + \theta_e \tau_e$ is not *that* negative²⁴. If θ cannot be so large, then $\tau_a < 0$. Otherwise, the sign of τ_a is ambiguous.

5.7 The effect of restructuring on power plant productivity

Empirically, the linear positive assumption does sufficiently restrict the magnitude of θ to conclude that restructuring *reduces* capacity-agnostic productivity. The effect of restructuring on productivity is between -4.53% and -0.46% across the two control group specifications using the 10% hypothesis interval. To maintain the same output as before, a plant would need to burn 0.46% to 4.53% more fuel if it has the same labor and non-fuel expenditures. The point estimates of the bounds are between -2.87% and -1.12% .

The effect size is economically meaningful. Fuel costs about \$3.25 per million British thermal units (EIA data, 2003 dollars). Multiplying \$3.25 by 1% of average fuel use in 2003, tells us that a 1% increase in fuel use (from a base of average fuel use) in 2003 costs \$1,085,802 in 2003 dollars for the average power plant. 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

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

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

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

Restructuring is the immediate effect of restructuring (τ_x), Post Trend is the annual trend in the effect of restructuring ($\tau_{POST,x}$), Pre Trend is the difference in trend between restructured and regulated states before restructuring ($\tau_{PRE,x}$), and Market Size is the coefficient on log market size (κ_x). Control Group IO refers to the specification using only investor owned power plants where such power plants in non-restructured states are the control group (Control Group 1). Control Group MUNI refers to the specification using both plants and municipally-owned power plants in restructured states as the control group (Control Group 2).

gas power plants. Taking the low-end of the EPA’s Social Cost of Carbon (\$11 per metric ton of carbon dioxide), this costs society about \$195,030 for a natural gas power plant using an average amount of fuel.

The negative effect of restructuring could either be a result of the combination of the competitive effects of restructuring, the [Averch and Johnson \(1962\)](#) effect, and inefficiencies from disintegration dominating the incentive effects of restructuring, *or* the effect of the policy could be negative initially because it *temporarily disrupted* power plant operations as they adjusted to the new policy, see [Holmes, Levine, and Schmitz \(2012\)](#) for a theory of why disruptions can cause monopolists to innovate less than firms that face greater competition.

To investigate the second case, I study the slope of the post-restructuring trend (τ_{POST}). We should be cautious in interpreting the trend as entirely causal. 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. But the disruption story is worth considering as one explanation for the results despite the difficulty in measuring it. I estimate the annual trend in the effect of restructuring to be between 0.22% and 1.02% using investor-owned power plants as the control group and between -0.84% and 0.96% using municipally-owned power plants as the control group. The number of years until the policy effect is positive is (if $\tau_{POST} \geq 0$),

$$-1 \times \frac{\tau}{\tau_{POST}}. \quad (63)$$

Because the results using municipal plants as the control group allow for $\tau_{POST} \leq 0$, 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 data²⁵.

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

I also measure the effect of market size on productivity. Market size has a theoretically ambiguous effect. Larger markets both encourage entry, increasing competition and making the residual demand curve more elastic, *and* shift the demand curve outward. See the large

²⁵I form a lower bound on the time until the sign of the policy reverses by solving a fractional linear programming problem because both τ and τ_{POST} 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 form these bounds.

literature on how market size affects market structure: [Shaked and Sutton \(1987\)](#), [Sutton \(1991\)](#), [Bresnahan and Reiss \(1991\)](#), [Melitz \(2003\)](#), and [Syverson \(2004\)](#)²⁶.

I find larger markets *increase* productivity. 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. 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.

5.8 The effect of restructuring on aggregate productivity

I also study how restructuring affects 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 by changing the way in which output is allocated across power plants. 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.

I call the aggregate productivity measure I use “geometric aggregate productivity” (GAP)

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

Table 3: Effect of restructuring on plant capacity-agnostic productivity

Parameter	Control Group	LB (10%)	LB	UB	UB (10%)
Restructuring	MUNI	−4.53%	−2.87%	−2.44%	−0.98%
	IO	−2.01%	−1.44%	−1.12%	−0.46%
Market Size	MUNI	0.88%	2.33%	3.09%	4.38%
	IO	0.99%	2.31%	2.95%	4.20%
Post Trend	MUNI	−0.84%	−0.22%	0.32%	0.96%
	IO	0.22%	0.54%	0.68%	1.02%
Pre Trend	MUNI	0.09%	0.22%	0.36%	0.51%
	IO	−0.08%	−0.04%	0.08%	0.12%
Restructuring (MUNI)	MUNI	−0.54%	0.79%	1.45%	3.19%
Post Trend (MUNI)	MUNI	−0.22%	0.28%	0.77%	1.28%
Pre Trend (MUNI)	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% hypothesis 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). Restructuring is the immediate effect of restructuring (τ_x), Post Trend is the annual trend in the effect of restructuring ($\tau_{POST,x}$), Pre Trend is the difference in trend between restructured and regulated states before restructuring ($\tau_{PRE,x}$), and Market Size is the coefficient on log market size (κ_x). I also report the parameters giving the “effect” of restructuring on for municipally owned plants (β , β_{POST} , β_{PRE}). The effect of restructuring is measured as the effect on investor owned utilities relative to this.

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

$$o_{it} = \frac{Q_{it}}{\sum_{j \in \text{state}} Q_{jt}} \quad (64)$$

$$\text{GAP}_{\text{state},t} = \prod_{i \in \text{state}} A_{it}^{o_{it}}. \quad (65)$$

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:

$$\log \text{GAP} = \underbrace{\frac{1}{N} \sum_i a_i}_{\text{Average Productivity}} + \underbrace{\sum_i \left(a_i - \frac{1}{N} \sum_i a_i \right) \times \left(o_i - \frac{1}{N} \sum_i o_i \right)}_{\text{Reallocation term}}. \quad (66)$$

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,

$$x_{\text{state},t} = \tau_x \times \text{Restruct}_{\text{state},t} + \tau_{PRE,x} Y_{\text{state},t} + \tau_{POST,x} Y_{\text{state},t} \times \mathbf{1}(Y_{\text{state},t} > 0) \quad (67)$$

$$+ \text{Year fixed effect} + \text{State fixed effect} + u_{\text{state},t}, \quad (68)$$

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

Restructuring caused average log productivity in a state to fall between -5.50% and -1.88% for investor-owned utilities and by -5.59% and -0.92% for all power plants. Average state productivity fell like plant-level productivity fell. But among investor-owned power plants, the plants directly affected by the policy, the reallocation effect of the policy was *positive*, between 0.77% and 3.09% ²⁷. 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 positive sign of the reallocation effect gives evidence that the new market for electricity more efficiently allocated output across the investor-owned power plants even though it reduced average productivity.

But the net effect of the reallocation and average productivity is negative. The effect of restructuring on GAP for all power plants is between -4.02% and -0.21% (point estimate between -2.29% and -1.78%) and among only investor-owned power plants GAP is between -3.64% and 0.26% (point estimate between -2.17% and -1.44%).

²⁷The reallocation effect including all power plants has a positive point estimate for the lower bound, but the 10% hypothesis interval extends from -0.17% to 2.52% .

Table 4: Effect of restructuring on geometric aggregate capacity-agnostic productivity (GAP)

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

Notes. Columns “LB (X%)” and “UB (X%)” give the upper and lower limits of the $X\%$ hypothesis 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. Restructuring refers to the immediate effect of restructuring, Post Trend refers to the post restructuring trend in the effect of restructuring, and Pre Trend indicates the difference in trend between restructured and non-restructured before restructuring.

5.9 Welfare interpretations

I close the empirical section of this article with a caveat. These results are not a welfare analysis of the restructuring policy. The negative productivity effect of restructuring by itself is not evidence that restructuring had a negative effect on total surplus. It may be efficient for power plants to reduce their productivity because, in the model of Section 2, productivity is a factor of production bought at a cost.

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 productivity and the reduction in productivity after restructuring was *efficient*.

In addition, productivity is not the only outcome affected by restructuring. Restructuring also changed the electricity prices paid by consumers. The price effect of restructuring matters for determining the total welfare effect of restructuring. The physical productivity effect is only a part of the welfare effect.

6 Conclusion

Standard production function estimation methods assume productivity is exogenous or else that any investment in it is observed. I show that we can allow productivity to be endogenous and meaningfully partially identify policy-relevant statistics of the productivity distribution. I prove a comparative static result that holds in a general economic model 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. I apply the inference method to form bounds on the effect of restructuring in the electricity industry on power plant productivity.

More structure could be added to narrow the identified set. An advantage of the comparative static method of this article is that the result does not depend delicately on the model used, but on the fundamental economics of the plant's production problem. It is worth searching for similarly robust restrictions we could put on the data to narrow the bounds. The search for identification in the production function literature has been long. The best hope for a path forward is by using identification restrictions that are based on the basic economics of production.

References

- Akerberg, D. A., K. Caves, and G. Frazer (2015). Identification properties of recent production function estimators. *Econometrica* 83, 2411–2451.
- Aghion, P. and P. Howitt (1992). A model of growth through creative destruction. *Econometrica* 60(2), 323–351.
- Andrews, D. and G. Soares (2010). Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica* 78(1), 119–157.
- Averch, H. and L. Johnson (1962). Behavior of the firm under regulatory constraint. *American Economic Review* 52(5), 1052–1069.
- Bickel, P. and B. Kleijn (2012). The semiparametric bernstein-von mises theorem. *The Annals of Statistics* 40(1), 206–237.
- Borenstein, S. and J. Bushnell (2015). The u.s. electricity industry after 20 years of restructuring. *NBER Working Paper* (21113).
- Bresnahan, T. and P. Reiss (1991). Entry and competition in concentrated markets. *The Journal of Political Economy* 99(5), 977–1009.
- Bugni, F., I. Canay, and X. Shi (2016). Inference for subvectors and other functions of partially identified parameters in moment inequality models. *Quantitative Economics*.
- Bushnell, J. and C. Wolfram (2009). The guy at the controls: Labor quality and power plant efficiency. *International Differences in the Business Practices and Productivity of Firms* (3), 79–102.
- Chernozhukov, V., H. Hong, and E. Tamer (2007). Estimation and confidence regions for parameter sets in econometric models. *Econometrica* 75(5), 1243–1284.
- Chernozhukov, V., S. Lee, and A. Rosen (2013). Intersection bounds: estimation and inference. *Econometrica* 81(2), 667–737.
- De Loecker, J. (2011). Product differentiation, multiproduct firms, and estimating the impact of trade liberalization on productivity. *Econometrica* 79(5), 1407–1451.
- De Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics* 5(3), 1–21.
- De Loecker, J., J. Eeckhout, and G. Unger (2019). The rise of market power and the microeconomic implications. *Working paper*.
- De Loecker, J. and F. Warzynski (2012). Markups and firm-level export status. *American Economic Review* 102(6), 2437–71.
- Doraszelski, U. and J. Jaumandreu (2013). R&d and productivity: Estimating endogenous productivity. *The Review of Economic Studies* 80(4), 1338–1383.

- Esary, J., F. Proschan, and D. Walkup (1967). Association of random variables, with applications. *The Annals of Mathematical Statistics* 38(5), 1466–1474.
- Fabrizio, K., N. Rose, and C. Wolfram (2007). Do markets reduce costs? assessing the impact of regulatory restructuring on us electric generation efficiency. *American Economic Review* 97(4), 1250–1277.
- Flynn, Z., A. Gandhi, and J. Traina (2019). Measuring markups with production data. *Working paper*.
- Gandhi, A., S. Navarro, and D. Rivers (2019). On the identification of production functions: how heterogeneous is productivity? *Working Paper*.
- Griliches, Z. and D. Jorgenson (1967). The explanation of productivity change. *The Review of Economic Studies* 34(3), 249–283.
- Griliches, Z. and J. Mairesse (1995). Production functions: the search for identification. *NBER Working Paper Series* (5067).
- Holmes, T., D. Levine, and J. Schmitz (2012). Monopoly and the incentive to innovate when adoption involves switchover disruptions. *American Economic Journal: Microeconomics* 4(3), 1–33.
- Hopenhayn, H. (1992). Entry, exit, and firm dynamics in long run equilibrium. *Econometrica* 60, 1127–1150.
- Hsieh, Y.-W., X. Shi, and M. Shum (2017). Inference on estimators defined by mathematical programming. *Working Paper*.
- Jovanovic, B. (1982). Selection and the evolution of industry. *Econometrica* 50(3), 649–670.
- Kaido, H., F. Molinari, and J. Stoye (2016). Confidence intervals for projections of partially identified parameters. *Working paper*.
- Kline, B. and E. Tamer (2016). Bayesian inference in a class of partially identified models. *Quantitative Economics* 7, 329–366.
- Knittel, C. (2002). Alternative regulatory methods and firm efficiency: Stochastic frontier evidence from the u.s. electricity industry. *The Review of Economics and Statistics* 84(3), 530–540.
- Levinsohn, J. and A. Petrin (2003). Estimating production functions using inputs to control for unobservables. *The Review of Economic Studies* 70(2), 317–341.
- Manski, C. and J. Pepper (2000). Monotone instrumental variables, with an application to the returns to schooling. *Econometrica* 68(4), 997–1012.
- Marschak, J. and W. Andrews (1944). Random simultaneous 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.
- 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.
- Sutton, J. (1991). *Sunk Costs and Market Structure*. MIT Press.
- Syverson, C. (2004). Product substitutability and productivity dispersion. *The Review of Economics and Statistics* 86(2), 534–550.
- Syverson, C. (2011). What determines productivity? *Journal of Economic Literature* 49(2), 326–365.
- Topkis, D. (1978). Minimizing a submodular function on a lattice. *Operations Research* 26(2), 305–321.
- Van Biesebroek, J. (2003). The effect of technology choice on automobile assembly plant productivity. *The Review of Economic Studies* 70(1), 167–198.
- Vives, X. (2008). Innovation and competitive pressure. *The Journal of Industrial Economics* 56(3), 419–469.

A Proofs

A.1 Proof of Lemma 1

Proof of Lemma 1.

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

$$\frac{\partial C}{\partial Q} = \frac{1}{A} \times \lambda, \quad (69)$$

where λ is the Lagrange multiplier on the output constraint.

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

$$\frac{\partial C}{\partial K} = -\lambda F_K(Z, K) = -A \times \frac{\partial C}{\partial Q} \times F_K(Z, K) \quad (70)$$

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

$$\frac{\partial C}{\partial K \partial Q} = -A \times \frac{\partial^2 C}{\partial Q^2} \times F_K(Z, K) - A \times \frac{\partial C}{\partial Q} \times F_{ZK}^\top \frac{\partial Z}{\partial Q} \quad (71)$$

By Assumption 2:

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

Therefore,

$$\frac{\partial C}{\partial K \partial Q} \leq 0. \quad (72)$$

□

A.2 Proof of Theorem 1

Proof of Theorem 1.

Assume there is an interior solution (Q^*) to the revenue maximization problem. The first order condition then requires,

$$\text{MR}(Q^*) = 0. \quad (73)$$

The second order condition requires that,

$$\text{MR}'(Q^*) \leq 0. \quad (74)$$

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

$$\text{MR}'(Q^{star}) \times Q + \text{MR}(Q) < 0, \quad (75)$$

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

A.3 Proof of Theorem 2

Proof of Theorem 2.

First, I prove the result when K is capacity. To establish the comparative static result, I make use of the [Topkis \(1978\)](#). The theorem requires that the constraint set is a lattice (closed under minimization and maximization operations), that the objective function is supermodular in all the choice variables, and that it has increasing differences in any given

choice with any parameter that we are trying to show the choice is increasing in. What this amounts to is showing that the objective function has positive pairwise cross partials with respect to any two choices and any one choice and any one parameter.

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

I write the optimization problem as,

$$\max_A \max_{K, \tilde{Q}} \sum_{t=1}^T \beta_t \times \left[\underbrace{P_t(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t}_{\text{Revenue}} - \underbrace{C_t(\tilde{Q}_t, W_t)}_{\text{Production costs}} - \underbrace{M_t(A_t, A_{t-1}, \eta_t)}_{\text{Technology cost}} - \underbrace{G_t(K_{t+1}, K_t, \kappa_t)}_{\text{Capital adjustment costs}} \right] \quad (76)$$

$$\text{ST: } \tilde{Q}_t A_t \leq K_t \quad (77)$$

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

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

$$\beta_t \times \left[P'(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t^2 + P(\tilde{Q}_t A_t, \xi_t) A_t - \frac{\partial C}{\partial \tilde{Q}_t}(\tilde{Q}_t, W_t) \right] \quad (78)$$

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

The derivative with respect to A_t is,

$$\beta_t \times \left[P''(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t^2 A_t^2 + 2P'(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t + P'(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t + P(\tilde{Q}_t A_t, \xi_t) \right] \quad (79)$$

$$= \beta_t \frac{\partial}{\partial Q_t} [\text{MR}(Q_t, \xi_t) Q_t] \geq 0, \quad (80)$$

by Assumption 4.

The derivative with respect to ξ_t is,

$$\beta_t \times A_t \frac{\partial}{\partial \xi_t} \left[P'(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t + P(\tilde{Q}_t A_t, \xi_t) \right] = \beta_t A_t \frac{\partial}{\partial \xi_t} \text{MR}(Q_t, \xi_t) \geq 0, \quad (81)$$

by Assumption 1.

The derivative with respect to $-W_t$ is,

$$\beta_t \times \frac{\partial C}{\partial \tilde{Q}_t \partial W_t} \geq 0, \quad (82)$$

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

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

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

$$-\beta_t \times \frac{\partial G_t}{\partial K_t}(K_{t+1}, K_t, \kappa_t) - \beta_{t-1} \frac{\partial G_{t-1}}{\partial K_t}(K_t, K_{t-1}, \kappa_{t-1}) \quad (83)$$

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

$$-\beta_t \frac{\partial G_t}{\partial K_{t+1} \partial K_t} \geq 0, \quad (84)$$

by Assumption 3.

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

$$-\beta_{t-1} \frac{\partial G_{t-1}}{\partial K_{t-1} \partial K_t} \geq 0, \quad (85)$$

by Assumption 3.

The derivative with respect to κ_t is,

$$-\beta_t \times \frac{\partial G_t}{\partial K_t \partial \kappa_t} \geq 0, \quad (86)$$

by Assumption 1.

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

$$-\beta_{t-1} \times \frac{\partial G_{t-1}}{\partial K_t \partial \kappa_{t-1}} \geq 0, \quad (87)$$

by Assumption 1.

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

By the [Topkis \(1978\)](#) theorem, $\tilde{Q}_t(s, A^T, W^T)$ and $K_t(s, A^T, W^T)$ are increasing in (s, A^T, W^T) .

Now, we move to the outer optimization problem. My goal here is to show that $A_t(s, -W^T)$ is increasing in $(s, -W^T)$. Let the Lagrange multiplier on the period t capacity restriction be λ_t .

By the envelope theorem, the derivative with respect to A_t is,

$$\beta_t \times \left[P'(\tilde{Q}_t A_t, \xi_t) \times \tilde{Q}_t^2 A_t + P(\tilde{Q}_t A_t, \xi_t) \times \tilde{Q}_t - \frac{\partial M_t}{\partial A_t}(A_t, A_{t-1}, \eta_t) \right] \quad (88)$$

$$- \beta_{t+1} \times \frac{\partial M_{t+1}}{\partial A_t}(A_{t+1}, A_t, \eta_{t+1}) - \lambda_t \tilde{Q}_t \quad (89)$$

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

$$\beta_t \times [P' \times \tilde{Q}_t A_t^2 + P \times A_t - C_{\tilde{Q}}] = \lambda_t A_t \quad (90)$$

$$\implies \beta_t \times [Q_t \text{MR}(Q_t, \xi_t) - \tilde{Q}_t \times C_{\tilde{Q}}] = \tilde{Q}_t A_t \lambda_t \quad (91)$$

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

$$\beta_t \left[\frac{\tilde{Q}_t}{A_t} \times \frac{\partial C}{\partial \tilde{Q}_t}(\tilde{Q}_t, W_t) - \frac{\partial M_t}{\partial A_t}(A_t, A_{t-1}, \eta_t) \right] \quad (92)$$

$$- \beta_{t+1} \times \frac{\partial M_{t+1}}{\partial A_t}(A_{t+1}, A_t, \eta_{t+1}) \quad (93)$$

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

$$\beta_t \left[\frac{\frac{\partial \tilde{Q}_t}{\partial A_{t+1}}}{A_t} \times \frac{\partial C}{\partial \tilde{Q}_t}(\tilde{Q}_t, W_t) + \frac{\tilde{Q}_t}{A_t} \times \frac{\partial^2 C}{\partial \tilde{Q}_t^2}(\tilde{Q}_t, W_t) \times \frac{\partial \tilde{Q}_t}{\partial A_{t+1}} \right] \quad (94)$$

$$- \beta_{t+1} \frac{\partial M}{\partial A_{t+1} \partial A_t}(A_{t+1}, A_t, \eta_t) \quad (95)$$

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

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

The derivative with respect to η_t is,

$$- \beta_t \times \frac{\partial M_t}{\partial A_t \partial \eta_t} \geq 0, \quad (96)$$

by Assumption 1.

Symmetrically, the derivative with respect to η_{t+1} is also positive.

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

The derivative of the interior objective with respect to ξ_t is,

$$\beta_t \frac{\partial P}{\partial \xi} \times \tilde{Q}_t A_t. \quad (97)$$

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

$$\beta_t \times \left[\frac{\partial P}{\partial Q \partial \xi} \times \frac{\partial Q_t}{\partial A_{t'}} \times Q_t + \frac{\partial P}{\partial \xi} \times \frac{\partial Q_t}{\partial A_{t'}} \right] = \quad (98)$$

$$\beta_t \times \frac{\partial Q_t}{\partial A_{t'}} \times \left[\frac{\partial P}{\partial Q \partial \xi} \times Q_t + \frac{\partial P}{\partial \xi} \right] = \quad (99)$$

$$\beta_t \times \frac{\partial Q_t}{\partial A_{t'}} \times \frac{\partial}{\partial \xi} \text{MR}(Q, \xi) \geq 0, \quad (100)$$

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

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

$$\beta_t \frac{\partial C}{\partial W_t} (\tilde{Q}_t, W_t) \quad (101)$$

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

$$\beta_t \frac{\partial C}{\partial \tilde{Q}_t \partial W_t} (\tilde{Q}_t, W_t) \times \frac{\partial \tilde{Q}_t}{\partial A_{t'}} \geq 0, \quad (102)$$

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

The derivative of the interior objective with respect to κ_t is,

$$-\beta_t \times \frac{\partial G_t}{\partial \kappa_t} \quad (103)$$

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

$$-\beta_t \times \left[\frac{\partial G_t}{\partial K_{t+1} \partial \kappa_t} \times \frac{\partial K_{t+1}}{\partial A_{t'}} + \frac{\partial G_t}{\partial K_t \partial \kappa_t} \times \frac{\partial K_t}{\partial A_{t'}} \right] \geq 0, \quad (104)$$

by Assumption 1 and from the above result that K_t is increasing in productivity.

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

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

$$\max_{K, A, \tilde{Q}} \sum_{t=1}^T \beta_t \times \left[P(\tilde{Q}_t A_t, \xi_t) \tilde{Q}_t A_t - C(\tilde{Q}_t, K_t, W_t) - M(A_t, A_{t-1}, \eta_t) - G_t(K_{t+1}, K_t) \right], \quad (105)$$

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

Derivative with respect to \tilde{Q}_t . The derivative with respect to \tilde{Q}_t is the same as in the previous proof. It is increasing in A_t and in s as shown above. It is also increasing in K_t because,

$$-\beta_t \frac{\partial C}{\partial K_t \partial \tilde{Q}_t} \geq 0, \quad (106)$$

by Assumption 2.

Derivative with respect to K_t . The derivative with respect to K_t is the same as in the previous proof except that $-\beta_t C_K$ is included as well. It is increasing in (A_t, K_{t-1}, K_{t+1}) and in s by the above argument and because, as shown above, it is increasing in \tilde{Q} .

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

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

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

The derivative with respect to ξ_t is,

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

by Assumption 1.

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

$$-\beta_t \frac{\partial M}{\partial A_{t-1} \partial A_{t-1}}(A_t, A_{t-1}, \eta_t) \geq 0, \quad (109)$$

by Assumption 3.

The derivative with respect to η_t is,

$$-\beta_t \frac{\partial M}{\partial \eta_t \partial A_t} \geq 0, \quad (110)$$

by Assumption 1.

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

$$-\beta_{t+1} \frac{\partial M}{\partial A_{t+1} \partial A_t} (A_{t+1}, A_t, \eta_{t+1}) \geq 0 \quad (111)$$

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

$$-\beta_{t+1} \frac{\partial M}{\partial \eta_{t+1} \partial A_t} M (A_{t+1}, A_t, \eta_{t+1}) \geq 0, \quad (112)$$

by Assumption 1.

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

A.4 Proof of Lemma 2

Proof of Lemma 2.

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

A.5 Proof of Theorem 3

Proof of Theorem 3.

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

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

By induction, \tilde{s}_t is positively associated for all t . $\tilde{s} = \cup_{t=0}^T \tilde{s}_t$. Because $(\tilde{s}_0, \tilde{s}_1)$ is positively associated, it follows that $(\tilde{s}_2, \tilde{s}_1, \tilde{s}_0)$ is positively associated because \tilde{s}_2 is an increasing function of $(\tilde{s}_1, \tilde{s}_0, v_2)$ which is positively associated. Following the induction argument, \tilde{s} itself is positively associated. □

A.6 Proof of Theorem 5

Proof of Theorem 5.

The derivative of f^α is,

$$Df^\alpha = (1 - \alpha) Df^0 + \alpha Df^1 \geq 0, \quad (113)$$

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

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

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

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

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

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

A.7 Proof of Theorem 6

Proof of Theorem 6.

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

Because $(q - \underline{q})$ 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 $\tilde{\Theta}_{\text{CD,LPA}}$.

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

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

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

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

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

We want to show that θ satisfies,

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

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

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

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

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

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

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

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

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

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

$$\text{cov} \left(\left[q - \theta_1 - z^\top \theta_z \right] \varphi_{1,k} (k - \underline{k}), \varphi_{2,q} (q - \underline{q}) \right) \quad (126)$$

$$+ \text{cov} \left(\left[q - \theta_1 - z^\top \theta_z \right] \varphi_{1,q} (q - \underline{q}), \varphi_{2,q} (q - \underline{q}) \right) \quad (127)$$

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

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

$$\text{cov} \left(\left[q - \theta_1 - z^\top \theta_z \right] \left(\varphi_{1,k} (k - \underline{k}) + \varphi_{1,q} (q - \underline{q}) \right), \varphi_{2,q} (q - \underline{q}) \right) \quad (129)$$

$$+ \text{cov} \left(\left[q - \theta_1 - z^\top \theta_z \right] \left(\varphi_{1,k} (k - \underline{k}) + \varphi_{1,q} (q - \underline{q}) \right), \varphi_{2,k} (k - \underline{k}) \right) \quad (130)$$

$$= \text{cov} \left(\left[q - \theta_1 - z^\top \theta_z \right] \left(\varphi_{1,k} (k - \underline{k}) + \varphi_{1,q} (q - \underline{q}) \right), \right. \quad (131)$$

$$\left. \varphi_{2,q} (q - \underline{q}) + \varphi_{2,k} (k - \underline{k}) \right) \geq 0 \quad (132)$$

For any function $\varphi_1 + \varphi_q q + \varphi_k k$ such that $\varphi_q \geq 0$ and $\varphi_k \geq 0$, the minimum value of the function across observed (q, k) is no smaller than $\varphi_1 + \varphi_q \underline{q} + \varphi_k \underline{k}$. Because the function must be positive, $\varphi_1 \geq -\varphi_q \underline{q} - \varphi_k \underline{k}$. 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 $\text{cov}(X_1, X_2 + \text{constant}) = \text{cov}(X_1, X_2)$. So the covariance restrictions hold for $\varphi_{2,1} + \varphi_{2,q} q + \varphi_{2,k} k$.

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

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

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

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

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

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

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

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

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

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

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