NBER WORKING PAPER SERIES

ESTIMATING PATH DEPENDENCE IN ENERGY TRANSITIONS

Kyle C. Meng

Working Paper 22536
http://www.nber.org/papers/w22536

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
August 2016

I thank Dan Benjamin, Javier Birchenall, Hoyt Bleakley, Chris Costello, Melissa Dell, Jonathan Dingel, Walker Hanlon, Akshaya Jha, Per Krusell, Derek Lemoine, Gary Libecap, Suresh Naidu, Trevor O'Grady, Maxim Pinkovskiy, Bernard Salanié, Steve Salant, Anna Thompsett, Inge van den Bijgaart, Tom Vogl, Sevgi Yuksel and seminar participants at the NBER DAE and EEE Summer Institutes, PERC, University of Gothenburg, and University of San Francisco for helpful comments. I am grateful to Kira Fabrizio, Josh Lewis, Jessika Trancik, and Catherine Wolfram for sharing data. Kayleigh Campbell Bierman and Andy Bilich provided excellent research assistance. Errors are my own. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Path Dependence in Energy Transitions
Kyle C. Meng
NBER Working Paper No. 22536
August 2016
JEL No. N51,N52,O41,Q35,Q43,Q54,Q58

ABSTRACT

Addressing climate change requires transitioning away from coal-based energy. Recent structural change models demonstrate that temporary interventions could induce permanent fuel switching when transitional dynamics exhibit strong path dependence. Exploiting changes in local coal supply driven by subsurface coal accessibility, I find that transitory shocks have strengthening effects on the fuel composition of two subsequent generations of U.S. electricity capital. To facilitate a structural interpretation, I develop a model which informs: tests that find scale effects as the relevant mechanism; recovery of the elasticity of substitution between coal and non-coal electricity; and simulations of future carbon emissions following temporary interventions.

Kyle C. Meng
Bren School of Environmental
Science and Management
Department of Economics
University of California, Santa Barbara
4416 Bren Hall
Santa Barbara, CA 93106
and NBER
kmeng@bren.ucsb.edu

1 Introduction

Economies tend to consume more of an abundant resource, turning to alternatives only after supply has sufficiently fallen (Herfindahl, 1967). However, when consumption of that resource entails externalities, it may be desirable to switch earlier. Coal is the most abundant and dirty fossil fuel. Local pollution from coal burning has long been a textbook case of social costs (Coase, 1960).[1] Globally, half of the cumulative anthropogenic carbon dioxide emitted since the Industrial Revolution can be traced to coal consumption (Boden and Andres, 2013). As a consequence, it is widely recognized that addressing environmental challenges such as climate change requires that the global economy undertakes a permanent structural transition away from coal-based energy.

How can an economy permanently switch out of a relatively abundant resource? Economic theory offers two views. When productivity is exogenous, any sustained transition towards a less abundant resource requires a *permanent* price intervention to offset relative supply. In contrast, recent models featuring endogenous structural change demonstrate that under sufficiently strong transitional dynamics, a large but *temporary* intervention may overcome supply forces, triggering long-term switching even after the intervention is lifted (Acemoglu et al., 2012, 2016). Economies that feature such dynamics are broadly characterized by strong path dependence,[2] or when the lagged effects of transitory shocks increase over time.

The presence of strong path dependence in the energy sector allows temporary policy interventions to induce a permanent switch away from coal. Temporary interventions are particularly advantageous for stock externalities, such as carbon emissions, if policy commitments over a long time horizon are not politically feasible or credible. Furthermore, strong path dependence may imply that recent breakthroughs in natural gas extraction can sustain fuel switching even if economic conditions change. Unfortunately, it remains unknown whether such transitional dynamics exist, as there is little empirical evidence on long-run path dependence in the energy sector.

This paper provides the first causal estimates of path dependence in the fuel composition of electricity capital, or capacity.[3] Examining Midwestern U.S. power plants over the 20th century, I find that local coal price shocks have lagged effects that increase in magnitude over two subsequent generations of coal capacity relative to non-coal capacity. Additional empirical tests find support for increasing returns to scale in electricity production as the relevant mechanism. Interpreted through a model of fuel switching in electricity production, I show how my reduced-form estimates map onto a formal definition of strong path dependence in which permanent transition is possible following a temporary shock.

Causal estimates of path dependence require variation in lagged county delivered coal prices that are uncorrelated with lagged and contemporaneous determinants of relative coal capacity, the ratio of coal to non-coal capacity. This ensures that lagged coefficients have a path dependence interpretation as distinguished more generally from persistence, or the ongoing effects of a time-invariant determinant (see Bleakley and Lin (2012)). Furthermore, because electricity capital is highly durable, detecting long-term effects on new capital investments requires county-level capacity and coal price data spanning much of the 20th century. Unfortunately, an extensive search of historical records revealed that such data was either never historically collected or no longer exists today.

[1] Recent empirical studies have shown substantial social costs due to coal-based local pollution (Chay and Greenstone, 2003, 2005; Barreca, Clay and Tarr, 2014; Clay, Lewis and Severnini, 2016; Beach and Hanlon, 2016; Hanlon, 2016).

[2] A related tradition in the economics of technological change has argued that path dependence may differentially lock-in certain technologies over time (Schumpeter, 1942; Schmookler, 1966; David, 1985; Aghion and Howitt, 1992; Arthur, 1994).

[3] Capacity is the maximum amount of electricity that can be produced (in watts) by that generator. It is typically used to describe the size of a electricity generator and is a measure of electricity capital stock.

To meet these requirements, this paper uses recently digitized spatial data on the location and depth of U.S. coal resources to construct local coal supply shocks driven by the interaction of time-invariant subsurface coal resources and time-varying aggregate mining technology. Specifically, I exploit the introduction of mechanized coal mining at the start of the 20th century which opened mines over previously inaccessible deep coal resources. I follow the logic of Hotelling's location model (Hotelling, 1929; Gaudet, Moreaux and Salant, 2001), in which delivered coal prices are approximated by distance to the nearest mine with variation occurring when mines open. Mechanized extraction allowed deep coal mining and altered the spatial distribution of delivered coal prices, making some previous price-setting shallow mines obsolete. As long as the decision to open a deep mine is driven by aggregate mining technology and not local coal demand, this distance-based proxy for delivered coal prices should satisfy my identifying assumptions.

Availability of mine openings since 1890 allows construction of a distance-based price proxy covering the entire 20th century. To obtain county fuel-specific capacity over the same period, I turn to modern power plant records containing the start and closing years of retired electricity generators. For both datasets, I employ available historical records to verify assumptions behind data construction and confirm data values. Using an event study specification, I examine whether distance to the obsolete shallow mine exhibits lagged effects on a county's relative coal capacity following the opening of a more proximate deep mine. To ensure these effects have a path dependence interpretation, my panel estimator, with counties experiencing the event in different times, includes contemporaneous coal prices, state-year fixed effects, and county fixed effects to control for time-varying and invariant determinants of relative coal capacity.

I find that distance to the obsolete mine has an increasingly negative effect on relative coal capacity for up to ten decades after becoming obsolete with notable jumps three and seven decades later of -3% and -5%, respectively. Importantly, these jumps coincide with the expected timing of two subsequent generations of new electricity capital. These estimates are robust to different model specifications, functional form assumptions, estimation samples, and various data construction and modeling choices. I also show that distance to an obsolete mine is uncorrelated with pre-trends in observable county characteristics that may directly affect relative coal capacity.

Path dependence in energy transitions following fuel price shocks can emerge from various mechanisms, operating at different scales. Prior literature offers two candidate mechanisms: local increasing returns to scale in electricity production (Nerlove, 1963; Christensen and Greene, 1976) and the accumulation of fuel-specific productivity (Acemoglu et al., 2012, 2016). To examine these competing mechanisms, I introduce a theoretical framework nesting (i) vintaged fuel-specific electricity capacity with fixed factor productivities (Komiya, 1962; Joskow, 1985, 1987; Atkeson and Kehoe, 1999), (ii) increasing returns to scale across all vintages of fuel-specific capacity (Nerlove, 1963; Christensen and Greene, 1976), and (iii) imperfect substitutability between coal and non-coal electricity (Acemoglu et al., 2012). Using predictions from this framework, I find empirical support for the presence of scale over productivity accumulation effects as the relevant mechanism. Additional empirical tests fail to detect other mechanisms such as scale effects in the residential and manufacturing sectors that consume electricity (Kline and Moretti, 2014; Severnini, 2014) and in the coal transport sector (Redding, Sturm and Wolf, 2011).

Isolating the relevant mechanism for path dependences allows my reduced-form estimates to be interpreted structurally within my modeling framework. Formally, strong path dependence occurs when there is either sufficiently large increasing returns to scale or a sufficiently high degree of substitutability between coal and non-coal electricity. Using this definition, my reduced-form estimates imply a value of 5 for the long-run elasticity of substitution between coal and non-coal electricity, a key parameter that is common across a broad class of structural change models driven by supply-side forces, first posited by Baumol (1967).[4] By recovering this parameter, this paper informs upon energy transitions due to other drivers of structural change. As an example, within the context of Acemoglu et al. (2012)'s model of optimal climate policy under directed technical change, my estimated elasticity implies that a temporary policy intervention would be sufficient to avoid climate disaster.

Strong path dependence suggests that it is possible for a temporary price shock to induce permanent fuel switching. The magnitude and/or duration of the required shock, however, depends on baseline relative coal prices. If baseline prices are low, a large or long-lasting price shock is needed to prevent the forces of path dependence from once again favoring coal after the shock. To analyze how temporary policies could overcome baseline coal prices at the national level, I simulate future U.S. electricity sector CO_2 emissions following temporary shocks of varying magnitude and duration. Simulations show that a shock at least 1.5 times larger than that of recent natural gas prices and lasting at least 3 decades is needed to induce a long-term emissions decline.

This paper is linked with several areas of empirical research. An emerging literature examines the empirics of short-run fuel switching under fixed capital stock. Aghion et al. (2016) explore firm-level patent differences in clean and dirty automobile technology across 80 countries in response to annual oil prices from 1986-2005, finding evidence of path dependence by demonstrating that a firm's current patenting responds to its stock of patents. Knittel, Metaxoglou and Trindade (2015) examine how coal to natural gas fuel-switching from recent contemporaneous price shocks may differ depending on the ownership and market structure of the electricity producer. This paper has the advantage of exploring transition dynamics over a time horizon long enough to detect changes in the fuel composition of highly durable electricity capital. This is particularly important for climate policy as it is unlikely that sufficient carbon emission abatement can be achieved without compositional changes to the existing energy capital stock.

In terms of identification strategy, this paper connects with a growing empirical literature examining dynamic effects following plausibly exogenous technological change that affects either the supply or productivity of a natural resource. This literature is generally classified by whether the technological advance puts a resource into or out of productive use. The adjustment literature explores the lagged effects following a persistent change in resource access (Feyrer, 2009; Nunn and Qian, 2011; Hornbeck and Keskin, 2014). The path dependence literature, which this paper follows, explores the lagged effects following the obsolescence of a previously productive or accessible natural resource (Bleakley and Lin, 2012; Dell, 2012; Hornbeck, 2012; Severnini, 2014; Hornbeck and Naidu, 2014; Glaeser, Kerr and Kerr, 2015).

The remainder of the paper is organized as follows. Section 2 provides motivating historical evidence on the U.S. electricity sector. Section 3 details the statistical challenges in identifying path dependence and a proposed solution. Section 4 covers data construction and verification checks. Section 5 presents reduced-form evidence of path dependence and robustness tests. Section 6 introduces a theoretical framework which informs empirical tests of potential mechanisms. Section 7 uses this framework to formally define path dependence strength, recover an elasticity of substitution, and simulate future CO_2 emissions pathways. Section 8 concludes the paper.

[4]Prominent models in this class consider structural change driven by intermediate sector-level differences in total factor productivity (Ngai and Pissarides, 2007), factor proportions and capital deepening (Acemoglu and Guerrieri, 2008), and directed technical change (Acemoglu, 2002; Acemoglu et al., 2012; Lemoine, 2016).

2 Background: Why are U.S. CO_2 emissions so persistently high?

The United States is the world's largest cumulative emitter of anthropogenic carbon dioxide (Boden and Andres, 2013) with historically high annual emissions even after accounting for income. Figure A.1 plots country-level CO_2 emissions per capita against GDP per capita in 1960 and 2000 for all non-OPEC countries showing that the U.S. was a positive 2.4 and 1.8 σ outlier in 1960 and 2000 respectively relative to countries with similar income. A major reason is the U.S.'s heavy reliance on coal for electricity, the most carbon intensive of energy inputs,[5] with roughly 40% of electricity capacity primarily burning coal since the 1960s (Energy Information Administration, 2013). There are two common arguments for why coal consumption has been so pervasive in the U.S. electricity sector.

The endowment argument highlights that the U.S. has always had a large abundance of coal compared to other primary fuels of similar energy content. The time-invariant role of coal availability is supported by a first-look at county-level data. Figure 1 plots 1920s county-level log relative coal capacity against log relative coal capacity in the 1990s. Counties with more relative coal capacity in the 1920s tend to have more relative coal capacity in the 1990s, though 1920 values only explain 4% of the variation in 1990.

The path dependence argument posits that the pervasiveness of coal-fired electricity results in part from the rising importance of coal at the moment of electricity's initial introduction. Electricity is an intermediary energy carrier produced from various primary energy inputs. Introduced around the start of the 20th century (see Figure A.2), electricity's arrival coincided with a transitional moment in the U.S. fuel mix with coal replacing wood as the dominant primary fuel (see Figure A.3). Additional county-level evidence supports this view. Figure 2 displays the average trajectories of relative coal capacity over the 20th century for counties that are close to (between 0 and 50 miles) and far away (between 200 and 250 miles) from time-invariant coal resources in the Illinois Coal Basin. Counties that are closer to coal resources consistently have higher relative coal capacity compared to counties that are further away. However, the difference in coal use between these sets of counties grow dramatically over the 20th century suggesting that time-varying factors may play an important role.

Neither Figures A.3 nor 2, however, directly imply the presence of path dependence. At the national level, other aggregate forces may have jointly determined the rise of coal and electricity, making causal inference difficult. At the county-level, the time path of relative coal capacity for counties close to and far from time-invariant coal resources in Figure 2 could be explain by such resources having persistent and increasing contemporaneous effects over the 20th century, prohibiting a path dependence interpretation. As detailed next, identifying path dependence requires examining the lagged effects of exogenous historical coal access that subsequently becomes obsolete. The introduction of mechanical coal mining, which altered the accessibility of different subsurface coal resources over time, provides a possible solution.

3 Identification strategy

This section first details the statistical challenges of identifying path dependence in relative coal-fired electricity capacity. I then discuss how the introduction of mechanized mining a data setting that may overcome these challenges.

[5] According to the U.S. Energy Information Agency, bituminous coal, the most common type of coal for electricity, yields 206 lbs of CO_2 per million BTU. By contrast, the CO_2 emission factor for gasoline and natural gas is 157 and 117 lbs of CO_2 per million BTU respectively. Available: `http://www.eia.gov/tools/faqs/faq.cfm?id=73&t=11`

3.1 Requirements for identifying path dependence

I consider a simple reduced-form statistical model to clarify how path dependence can be identified. There is one location with coal and non-coal electricity subsectors, $j \in \{c, n\}$. $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$ is relative coal capacity in period t which depreciates completely between each period. w_{ct} is the delivered coal price. Relative coal capacity has the following data generating process:

$$\widetilde{X}_t = \rho \widetilde{X}_{t-1} + \pi w_{ct} + \mu_t$$

where π is the contemporaneous effect of coal price on relative coal capacity. ρ captures path dependence.[6] μ_t is a contemporaneous error term that includes both time-varying and time-invariant determinants of relative coal capacity. The presence of the latter, in particular, implies that the estimated auto-regressive coefficient from the model above could not distinguish between path dependence following a transitory shock and persistence due to ongoing determinants. Instead, consider rewriting with a one-period lag:

$$\widetilde{X}_t = \rho \pi w_{ct-1} + \pi w_{ct} + \xi_t$$

where $\xi_t = \rho^2 \widetilde{X}_{t-2} + \rho \mu_{t-1} + \mu_t$. Identifying path dependence requires $E[w_{ct-1}\xi_t | w_{ct}] = 0$ and $E[w_{ct}\xi_t | w_{ct-1}] = 0$. Identifying variation in both lagged and contemporaneous coal prices must be uncorrelated with lagged and contemporaneous determinants of relative coal capacity. In particular, to support a path dependence interpretation, lagged prices must be uncorrelated with contemporaneous determinants such that it is transitory or becomes obsolete. Notice this condition is similar in spirit to the exclusion restriction in a standard instrumental variables setup by requiring that past prices affect future relative capacity only through past relative capacity.

3.2 Proposed solution: variation in local coal supply from mechanized mining

To meet these identifying assumptions, I use local coal supply shocks driven by subsurface coal geology and aggregate mining technology to construct a proxy measure of local delivered coal prices that is plausibly exogenous to local coal demand. Specifically, I exploit the aggregate introduction of mechanized mining which allowed extraction over previously inaccessible deep coal resources. The procedure follows the price-setting logic of Hotelling (1929)'s canonical location model.

Access to deeper coal from mechanized mining Prior to the 20th century, most coal in the U.S. was manually extracted which limited the spatial extent of extractable coal primarily over shallow resources generally less than 200 feet from the surface (Fisher, 1910; Speight, 1994).[7] Manual mining made way for mechanized extraction around the turn of the century and eventually came to dominate coal mining. Figure A.4 shows that nearly all increases in bituminous coal mining, the type of coal most commonly burned for electricity, between 1890-1930 occurred as a result of mechanized extraction. Chief among the benefits of mechanization was the introduction of mechanized drills which allowed for excavation of deeper coal resources that was previously inaccessible. Using the logic discussed next, this change in aggregate mining technology, when interacted with the location and depth of coal resources, altered the spatial distribution of local delivered coal prices.

[6] This is true only under complete capital depreciation between each period, as assumed here. The coefficients estimated in Section 5 and interpreted structurally in Section 7 do not assume complete depreciation. Furthermore, ρ is a reduced-form parameter as it does not specify a particular mechanism, which will be considered in Section 6.

[7] In geology, resources refer to all known underground deposits while reserves are the subset of resources which can be economically mined using existing mining technologies.

Defining a proxy for delivered coal prices via Hotelling's location model The Hotelling location model examines how producers compete over a spatial distribution of consumers by choosing location and prices (Hotelling, 1929; C. d'Aspremont, 1979). Because coal is costly to ship across space, a variant of the Herfindahl Principle (Herfindahl, 1967) emerges in the competitive equilibrium with each county buying coal only from the nearest mine. The equilibrium delivered coal price faced by any consuming county is captured, inter alia, by physical distance to the nearest mine.[8]

The basis for my identification strategy employs a natural resource extension following Kolstad (1994) and Gaudet, Moreaux and Salant (2001) in which the characteristics of subsurface coal resources determines coal mine location in a plausibly exogenous manner. Beneath every location is a coal resource of varying depth. Coal mining requires overcoming a depth-dependent fixed cost. Advances in mining technology, such as mechanization, enable coal mining in locations with deeper resources. The resulting distance-based price proxy is therefore generated by the interaction between subsurface coal geology and aggregate mining technology.

Figure 3 illustrates how delivered coal prices and path dependence are captured in a 1-dimensional setting.[9] There are 3 locations with coal, A, B, and C, sequentially located on a line. Locations A and C have shallow resources while location B has a deep resource. Consumers are uniformly distributed between A and C. There are 2 time periods with complete capital depreciation in between. Consider the two counties i and i'. In the first period with only manual mining (left panel), both counties, being closer to A than C, purchase coal from the mine at A. Because county i is relatively closer to A than county i', it faces lower delivered coal prices and thus has higher coal capacity. In the second period, the introduction of mechanical extraction allows deep coal mining over location B (right panel). Because both counties are closer to B than A, both should purchase from mine B with mine A becoming obsolete to both counties.[10] In the absence of path dependence, coal capacity during period two should be captured by distance to B, the proxy for contemporaneous coal price. However, under path dependence, distance to the obsolete mine at A may still exert an influence on second period coal capacity, allowing more coal capacity in county i relative to county i' than distance to B alone would determine. If path dependence were particularly strong, county i may continue to have more coal capacity than county i' in the second period even if distance to mine A may no longer determine contemporaneous prices.

Hotelling's model defines my treatment of interest: distance to the obsolete coal mine in the periods after obsolescence. Importantly, to satisfy my identifying assumptions in Section 3.1, the decision to open a deep coal mine must be determined by existing mining technology and not by local coal demand. This informs how my distanced-based proxy is constructed, which is detailed next.

4 Data construction and verification

Estimating path dependence requires both county-level delivered coal prices and fuel-specific electricity capacity that span much of the 20th century. As summarized in Appendix Section B, an extensive search through historical records reveals that this data was either never collected or is no longer available. This section details how spatial data on coal resources and mines are used to construct a proxy measure of county

[8] Equilibrium delivered coal prices for any county is a function of distance to nearest mine, the resource scarcity rent at the nearest mine, and distance to other competing mines (Hotelling, 1929; C. d'Aspremont, 1979).

[9] See Gaudet, Moreaux and Salant (2001) for a formal examination of the general case in 2-dimensional space for any number of arbitrarily located counties and coal producers. Implications are the same for the more general case.

[10] More generally, obsolescence is defined for a county-mine pair. A mine may become obsolete for one county but may still be the supplier for another county.

delivered coal prices over the 20th century along with verification tests for the Hotelling model that enables this construction. I then present how a county panel of fuel-specific electricity capacity over the same period is reconstructed using modern power plant records followed by verification against available more aggregated historical data. Appendix Section A details all data used.

4.1 Constructing a distance-based proxy for county delivered coal prices

U.S.G.S. assessment of coal resources and mines in the Illinois Basin I use spatial data from the National Coal Resource Assessment (N.C.R.A.) recently digitized by the United States Geological Survey (U.S.G.S.) (East, 2012). The N.C.R.A. includes GIS shape files for coal resources at discrete depths from the surface for all major U.S. coal basins (see Figure A.5).

Of these basins, I consider only the Illinois Basin because it has four features that reasonably approximate the Hotelling model in Section 3.2. First, the Illinois Basin has a relatively simple dish-like shape with shallow coal in the outer regions that deepen towards the center of the basin. Figure 4 maps the location of coal resources by depth using N.C.R.A. data. The lighter shaded areas show coal resources that are less than 200 feet below the surface and approximate resources that were recoverable when only manual mining was available. The darker shaded areas in Figure 4 indicate the location of coal resources situated at depths greater than 200 feet deep. More geologically complicated basins such as the Appalachian Basin often have overlapping coal resources of different depths under the same location. When coal resources are overlapping, mining of a shallow resource could directly lead to mining of deeper resources, possibly violating the exogeneity of deep mine openings.

Second, coal quality is relatively homogeneous in the Illinois Basin compared with other eastern U.S. coal, reducing concerns that path dependence could occur because electricity generators are being built to burn local grades of coal, as noted by Joskow (1985, 1987). The three maps shown in Figure A.6 indicate the heat, sulfur, and ash content of coal produced across the U.S. during 1990-1999 obtained from EIA 423 coal procurement data. Compared with coal found in the Appalachian Basin, Illinois Basin coal exhibits less variation in heat, sulfur and ash content. Third, as Figure A.6 shows, the size of counties around the Illinois Basin are also relatively uniform compared to other parts of the country. This means that measures of straight-line distance between mine location and the area-weighted centroid of a county would be less influenced by variation in county size. Fourth, the subsurface structure of the Illinois basin has been characterized since the start of the 20th century such that one can credibly interpret the coal resources shown in the N.C.R.A. database as representing resources known throughout the 20th century.[11]

Mine data availability from the N.C.R.A. also favors the Illinois Basin. Of the coal basins surveyed for the N.C.R.A, only the Illinois Basin assessment includes geolocated coal mines with opening and closing dates since 1880.[12] I restrict the sample to just mines over large coal resources as they are more likely to meet the assumption for exogenous delivered coal prices discussed in Section 3.2. Larger coal resources have lower scarcity rents when mined. This implies that the mining decision is more likely to be determined by existing mining technology than by local coal demand. Furthermore, for this mines, the resulting delivered coal prices is more likely to be dominated by transport costs than by scarcity rents. Figure 4 shows the location of all large coal mines that operated after 1880, defined as mines that are above the 95% percentile in spatial area.

[11]See, for example, the Illinois coal basin shown in Campbell (1908)'s map of U.S. coal fields as compared with that shown in Figure A.5.

[12]Historical coal mine data for other coal basins are typically held by state-level agencies. However, they are based on the "final map" for a coal mine which only notes the closing year.

Construction procedure Under the Hotelling model, the delivered coal price faced by a county is approximated by the straight-line distance to the nearest mine. A previous price-setting shallow mine becomes obsolete to a county when a more closely located deep mine opens. The county distance to this shallow mine after obsolescence is my treatment of interest.

Construction of this distance measure takes several steps. Using N.C.R.A. data, I first identify shallow and deep coal mines by spatially overlaying the coordinates of all Illinois Basin large mines that ever existed since 1890 onto shape files for the basin's shallow and deep coal resources. Using the opening and closing years of each mine, I then construct a mine-by-decade panel indicating when each mine, shallow or deep, was in operation.[13] Next, I restrict my sample to counties for which straight-line distance from the area-weighted centroid to the nearest Illinois coal resource is (i) less than that to the nearest Appalachian coal resource and (ii) less than 250 km. The first restriction ensures that sample counties are unlikely to be influenced by Appalachian coal. The second restriction ensures that sample counties are unlikely to be influenced by other coal basins. Results are not sensitive to either specific restriction, as shown in Section 5.4. The resulting benchmark county sample is shown in Figure A.7 along with the location of Illinois and Appalachian Basins.

Next, for each county and decade, I search for the nearest mine according to the straight-line distance between that county's area-weighted centroid and the mine mouth. This distance is my proxy measure for contemporaneous county delivered coal price. Figure A.8 maps county distance to nearest mine from 1890-1950. It is evident that the timing and location of mine openings has altered the spatial distribution of delivered coal prices in the sample over the 20th century.

The distance-based proxy for delivered coal prices shown in Figure A.8 is driven by the timing and location of all large mine openings. To obtain a more plausibly exogenous treatment, I isolate price variation occurring right before a county first switches from shallow to deep coal, making the previous price-setting shallow mine obsolete for that county. Because coal depths vary spatially, this event occurs at different moments for each county. Figure A.9 plots when this switch occurs for the counties in the sample, stacked arbitrarily on the y-axis. The light shaded areas show the period before and during while the nearest mine was shallow. The darker shaded areas show the period after a closer deep mine opens. Because many deep mines opened at the start of the 20th century, 40% of sample counties first experience the switch during the 1900s. Another 37% experiences the switch during the 1960s.

Verifying constructed data I consider three verification checks for my distance-based proxy of county delivered coal prices under the Hotelling model.

First, the Hotelling model assumes that delivered coal prices are dominated by transport costs. This is consistent with previous research noting the high costs of transporting coal, the heaviest primary fuel by heat content (Joskow, 1987; McNerney, Farmer and Trancik, 2011). Figure A.10 supports this assumption showing that transport costs were between 40-60% of national delivered coal prices for much of the first half of the 20th century, declining thereafter.

Second, the Hotelling model implies a variant of the Herfindahl Principle (Herfindahl, 1967) whereby counties buy coal only from the nearest mine. To test whether this implication is reasonable, I turn to U.S. Energy Information Agency (EIA) data on coal procurement during 1990-1999. The EIA 423 forms document the price and quantity of coal delivered to all power plants, but unfortunately only notes county of origin of the delivered coal and not the specific mine. Thus, the EIA 423 data allows an indirect test by examining whether coal delivered to a destination county varies according to straight-line distance between

[13] Specifically, if the mine was in operation for any year in a decade, I note that it was in operation during that decade.

the area-weighted centroids of destination and origin counties.[14] For every coal-buying county in the U.S., the left panel of Figure A.11 plots the share of total coal purchased and log delivered coal price averaged over 1990-1999 against the rank in bilateral distance between destination and origin counties. For a destination country, a rank of 1 means that the coal is sourced from within the same county, while a rank of 2 means the coal is sourced from the nearest neighboring county. There is a steep drop in coal shares during initial increases in distance rank which flattens thereafter. Delivered prices expectedly show a mirrored relationship. The right panel of Figure A.11 provide the same information but for the subset of sample counties.

For my final check, I regress observed county delivered coal prices available from EIA 423 forms during the 1970s, 1980s, and 1990s against my distance-based proxy. A tight fit is not expected nor necessarily desired as observed delivered prices contained many determinants in addition to local coal supply shocks driven by mine openings. However, a strong correlation lends assurance that my distance-based proxy captures variance in delivered coal prices. Table A.1 shows coefficients on log distance to nearest mine, my proxy measure, separately estimated for each decade. I detect a statistically significant correlation for all three decades regardless of whether the relationship is estimated over my sample counties or all counties with at least one coal-fired power plant.

4.2 Constructing county fuel-specific capacity

Modern EIA 860 forms on active and retired generators Because fuel-specific capacity at or below the county level is not available throughout the 20th century, I instead turn to generator-level data[15] contained in EIA 860 forms over 1990-2012 to construct a historical fuel-county-decade panel of electricity capacity from 1890-1990.[16] This is possible because the EIA 860 mandates reporting of both active and retired generators located on currently operating power plants. Reported generator variables relevant for this paper include capacity, opening and closing year, primary input fuel, thermal efficiency,[17], prime-mover,[18] and county of location.

Construction procedure follows two steps. First, I create a cross-sectional generator database taking the most recently reported data across all 1990-2012 EIA-860 forms to ensure that I include all available power plants. Next, using the opening and closing years of each generator, I construct a balanced fuel-county-decade panel of electricity capacity from 1890-1990.

Three assumptions are required in order for this constructed panel to match historical data. First, all power plants since 1890 must continue to have active generators today. If an entire power plant retires, their generators would not show up in the modern EIA 860 forms. Second, a generator must not change capacity during its lifetime. Third, a generator must not change its primary input fuel. Section 6.1 discusses engineering reasons for why a generator's capacity and input fuel are unlikely to change over time. Nonetheless, my constructed capacity data will violate these three assumptions to some degree. To examine the magnitude of such violations, I turn to a series of verification checks using available historical data at more aggregated levels.

[14] Because multiple mines exist in a given county, it is unlikely that any county buys coal exclusively from a single origin county, but one would expect more coal to be sourced from nearby counties.

[15] A power plant typically contains several generators of different vintages.

[16] I only include power plants owned by utility companies because units owned by non-utilities are inconsistently reported across EIA 860 forms.

[17] Thermal efficiency measures how efficiently input energy is converted into electricity. For a given generator, the thermal efficiency is the ratio of the heat content from electricity produced (in BTUs) over the heat content of fuel consumed (in BTUs). This was directly collected from 1990-1995.

[18] The prime-mover is the machine which converts fuel into electricity. The most common prime-movers in electricity generation are steam engines, hydropower, and the internal combustion machine.

Verifying constructed data I use available historical data at more aggregate levels to test my first assumption. In particular, capacity by prime-mover is available since 1920 at the national (U.S. Census Bureau, 1975) and power plant levels (Lewis, 2014). Because generators operate over multiple decades, my constructed data would be differentially biased depending on whether a power plant omitted from EIA 860 records retired prior to or after 1920. The former would imply a bias in my capacity values for much of the 20th century while the latter would imply bias in my new capacity values for each decade.

Panel (A) of Figure A.12 compares national steam-powered capacity during 1920-1970 from the U.S. Historical Census (U.S. Census Bureau, 1975) against national steam-powered capacity aggregated from my panel dataset constructed out of modern EIA 860 forms. Steam-powered capacity is typically fueled by coal, oil, or gas. There appears to be some underreporting in my reconstructed data from 1920-1955 with convergence between the two datasets afterwards. Panel (B) of Figure A.12 provides a similar national comparison but for new steam-power capacity. National new capacity from the constructed data closely track that from the historical census suggesting that I am likely observing power plants that have opened after 1920. Panels (C) and (D) of Figure A.12 provides similar national comparisons of cumulative and new capacity but for hydropower capacity. Again, my constructed data appear to be missing some cumulative hydropower capacity between 1920-1955 (Panel C) but the missing power plants are likely those built before 1920 (Panel D).

To provide a finer comparison, Figure A.13 replicates Figure A.12 but only for the eleven states covering the estimating sample as shown in Figure A.7 by aggregating power plant-level historical data available from Lewis (2014).[19] As with the national comparison, Panel (A) shows that my constructed data under reports cumulative steam-powered capacity over 1920-1955. New capacity shown in Panel (B) suggests that missing power plants are likely those built prior to 1920. A similar pattern is shown for hydropower capacity across the eleven sample states in Panels (C) and Panel (D). In summary, these comparisons suggest that my reconstructed data may be under reporting cumulative relative coal capacity but is likely accurately capturing new electricity capacity built since 1920.

Turning now to my two other assumptions, I examine whether generator capacity and fuel input have changed over time by exploring the consistency in reported values both across EIA 860 forms from 1990-2012 and by comparing the 2012 EIA 860 data against the earliest available generator-level dataset from 1980. Table A.2 examines the consistency of several generator variables within the 1990-2012 EIA 860 records. For each variable shown across columns of Table A.2, row values indicate the percentage of EIA 860 reports from 1990-2011 that differed from the value recorded in 2012. For example, 94% of generators reported using the same primary fuel throughout 1990-2011 as was reported in 2012. Likewise, for reported generator capacity, opening year, and retirement year, 75%, 97% and 80% of generators respectively reported the same value in 1990-2011 as was reported in 2012.

To further examine consistency of generator values going further back, I digitized the 1980 EIA "Inventory of Power Plants in the United States", the earliest available comprehensive generator-level dataset, collected in the late 1970s (see Section B for more details). Figure A.14 plots the generator-level capacity reported in 2012 against the capacity reported in the late 1970s. The relationship between the two reported values is nearly one-to-one. Table A.3 shows the distribution of reported primary fuel in 2012 conditional on primary fuel reported in the late 1970s. There has been very little fuel switching amongst coal-fired generators with

[19] While the historical power plant-level dataset constructed by Lewis (2014) includes power plant names, only fuzzing string matching with modern EIA-860 data can be conducted. It is impossible to verify whether any string mismatches is due to power plant name changes, data entry errors, or actual missing data in the EIA-860 forms. As such, comparing capacity at the state level provides a more straightforward test of data censuring.

92% of generators that reported using coal in the late 1970s reporting continued coal use in 2012. Fuel switching occurs slightly more frequently between natural gas and oil-fired generators with 77% and 75% of generators which reported use of natural gas and oil respectively in the late 1970s noting the same fuel use in 2012. There is no evidence of any fuel switching since the late 1970s for generators that use nuclear and hydro power. In summary, it appears reasonable to assume consistency in the capacity and input fuel of a generator over its lifetime.

5 Reduced-form results

This section presents reduced-form evidence of path dependence in relative coal capacity. To guide regression model selection, I first test whether pre-trends in various county characteristics are correlated with distance to the eventual obsolete coal mine across various specifications. I then estimate the effects of distance to the eventual obsolete mine before and after the event followed by various robustness checks.

5.1 Examining pre-trends in county characteristics

My distance-based proxy for local delivered coal prices must be uncorrelated with contemporaneous determinants of relative coal capacity. As an indirect test of this assumption, Table 1 examines whether distance to the eventual obsolete mine is correlated with county characteristics (Haines and Inter-university Consortium for Political and Social Research, 2010) when the mine was nearest to that county and thus the active price setter. These characteristics are indicated down each row of Table 1. I examine total and urban population. For the manufacturing sector, I further examine the number of establishments, number of wage earners per establishment, capital value, and total output.

Each column of Table 1 estimates a different model. For all models, standard errors are clustered at the county level, allowing for heteroscedasticity and arbitrary serial correlation within a county. In Column (1), I estimate the effect of log distance to the eventual obsolete mine on log characteristics. Columns (2) and (3) examine pre-trends by estimating first-differenced log characteristics, with Column (3) further including fixed effects. Columns (2) and (3) are analogous to a fixed-effects panel estimator in levels, with and without state-year fixed effects. Column (3) shows that pre-trends in county characteristics are uncorrelated with distance to the obsolete mine when time-invariant county unobservables and state-year fixed effects are removed.

5.2 Main specification

My benchmark model is further informed by two properties of county capacity data. First, the EIA 860 forms omit generators with capacity less than 1 MW. As a consequence, county-level values of relative coal capacity exhibit a large number of zero or indeterminate observations. Second, the distribution of relative coal capacity is left-skewed. To address the first concern, I add a value of 1 to both coal and non-coal capacity for all county-decade observations. This is reasonable as it is possible that most counties have some electricity capacity even if it is not reported in EIA 860 forms. One could address the second concern by applying a log transformation to the outcome variable. However, a log transformation is sensitive to low values which is now exacerbated by my solution to the first concern. Instead, I elect to estimate a Poisson fixed effects model as my benchmark.[20] Specifically, my benchmark reduced-form specification for modeling either relative coal capacity, or gross new relative coal capacity,[21] in county i, state s, decade t is:

$$\widetilde{X}_{ist} = exp\left(\sum_{\substack{-2\leq\tau\leq 10 \\ \tau\neq 0}} \beta^o_\tau[\ln dist^o_i \times \mathbf{1}(t=\tau)] + \sum_{\substack{-2\leq\tau\leq 10 \\ \tau\neq 0}} \gamma_\tau \mathbf{1}(t=\tau) + \beta^c \ln dist_{it} + \mu_i + \phi_{st} \right) + \epsilon_{ist} \quad (1)$$

where $\ln dist^o_i$ is natural log distance to the obsolete shallow mine. τ is the event time index , with $\tau = 0$ indicating the event period when the shallow mine was nearest to county i and thus the active price setter.[22] γ_τ are event time dummies. Following the pre-trend tests in Section 5.1, Equation 1 includes a vector of county fixed effects, μ_i and state-year fixed effects, ϕ_{st}. Additionally, I include a proxy for contemporaneous delivered coal price, the distance to current nearest mine, $\ln dist_{it}$, to allow for cleaner detection of path dependence. β^o_τ is the reduced-form parameter of interest. When $\tau > 0$, β^o_τ captures lagged effects of distance to the obsolete mine since obsolescence relative to its effect when the mine was active. Similarly, when $\tau < 0$, β^o_τ captures relative lead effects. Finally, standard Poisson models impose that the first and second moments of the outcome be equal. To address this issue, I estimate Eq. 1 with county-level clustered standard errors which relaxes this restriction by allowing arbitrary forms of heteroskedasticity and serial correlation within a county.

5.3 Main estimates of path dependence

Table 2 shows lead and lagged effects of distance to the obsolete mine, β^o_τ, and the contemporaneous effect of distance to the nearest mine, β^c, from Eq. 1. Column (1) models coal capacity. I detect lagged effects that are statistically significant and which become more negative in magnitude over the ten decades following obsolescence. I do not detect lead effects. Coal capacity in each period includes both capital depreciation and investment. To isolate effects on gross new capital investments, Column (2) examines gross new coal capacity, showing a similar pattern. Columns (3) and (4) shows the effect of distance to the obsolete mine on capacity and gross new capacity using all other fuels (i.e. oil, natural gas, nuclear, hydro, etc) respectively. I do not detect a systematic pattern in the lagged effects on non-coal capacity. Strangely, there appears to be a 1-period lead effect on non-coal capacity and a contemporaneous effect on new non-coal capacity.

Columns (5) and (6) display my main result: path dependence in relative coal capacity. Consistent with Columns (1) and (3), Column (5) shows distance to the obsolete mine has increasingly negative effects on relative coal capacity up to ten decades after obsolescence. I do not detect lead effects. These lead and lagged effects are plotted as a thick solid line in Figure 5. Crucially, I detect two distinct jumps after obsolescence. A 1% increase in distance to the obsolete mine lowers relative coal capacity by 3% and 5% three and seven decades later relative to the contemporaneous effect. Because the average duration of the period when the mine was active is 1.7 decades and sample generators on average last 4.7 decades (see lifespan distribution in Figure A.15), these jumps coincide roughly with the expected timing of two subsequent generations of new electricity capacity. This timing is confirmed by the pattern of lagged effects on gross new relative coal

[20] Additionally, the Poisson distribution benefits from being a member of the linear exponential family such that even if the density is misspecified, one can still obtain consistent point estimates through quasi-MLE provided that the conditional mean function is correctly specified.

[21] Gross new capacity captures gross investment and thus is not the same as the year-to-year difference in capacity which also includes depreciation.

[22] Because mine openings occur irregularly, a shallow mine may be the closest to a given county over several decades. Average duration of event period for sample counties is 1.7 decades.

capacity shown in Column (6). For a sense of magnitude, one can compare lagged and contemporaneous effects. Lagged effects three and seven decades later are three and four times stronger, respectively, than the contemporaneous effect.[23]

5.4 Robustness checks

I conduct a series of robustness checks to examine the stability of my benchmark estimate.

Nonlinearity Eq. 1 implicitly assumes that the relationship between relative coal capacity and distance to the obsolete mine is an isoelastic function, captured by a linear coefficient. While this is suitable for the theory to be considered in Section 6, it need not be empirically true. Figure 6 examines whether the data supports linearity by estimating a variant of Eq. 1 that breaks distance to obsolete mine into 50 mile-wide bins, allowing for a flexible relationship with relative coal capacity for each period after obsolescence. Figure 6 shows the estimated dummy coefficients one, three, five, seven, and nine decades after obsolescence. Linearity appears to be a reasonable assumption.

Model specification Table 3 examines alternative model specifications. In Column (1), I estimate a stripped-down version of Eq. 1 by including only county fixed effects and event time dummies. I detect statistically significant lagged effects two to five decades after obsolescence but they do not become more negative over time. The estimates in Column (1), however, may be biased towards zero as the model does not control for time-varying contemporaneous determinants of relative coal capacity which may be positively correlated with distance to the obsolete mine. Columns (2) and (3) incrementally includes distance to nearest mine, a direct proxy for contemporaneous delivered coal price, and state-year fixed effects state trends. Lagged coefficients become more negative and, in general, more precise as these additional controls are added. Column (3) replicates my benchmark estimate from Column (5) of Table 2. To address remaining exogeneity concerns, in Column (4) I additionally control for total population, number of manufacturing establishments, and manufacturing employment, three county covariates available over the entire 20th century. The magnitude of lagged effects are very similar to the benchmark estimate.

Estimating sample Table 4 examines different county sample definitions. For my benchmark model, reproduced as Column (1), the sample is restricted to counties within 250 miles of the nearest Illinois Basin coal and are situated closer to Illinois Basin coal than to Appalachian Basin coal. Recall that the rationale for both sample restrictions is to lessen the competing effects of coal resources in neighboring basins. In Column (2), I further weaken the effects of other coal basins by restricting my sample to counties within 200 miles of the nearest Illinois Basin coal resource. I find effects similar to my benchmark. In Column (3), I allow more counties into my sample by increasing the distance threshold to within 300 miles of the nearest Illinois Basin coal resource. In Column (4), I allow counties that are situated closer to the Appalachian Basin into the sample. For both larger samples, effect sizes are slightly smaller possibly due to the competing influence of other coal resources though coefficients are within the confidence interval of my benchmark estimate. In Column (5), I consider only counties that first faced obsolescence prior to 1960 to test whether there are differences based on when counties are treated. I am unable to estimate lead effects as my data set begins in 1890 and many counties were treated in 1900. However, the overall pattern of lagged coefficients is similar to the benchmark estimate.

[23] Three decades later: $\frac{-4.21-2.34}{-2.34} = 2.80$. Seven decades later: $\frac{-6.74-2.34}{-2.34} = 3.88$

Data construction choices Table 5 examines whether estimates are sensitive to how my distance-based proxy for delivered coal prices and relative coal capacity are constructed. Column (1) replicates the benchmark result. Under the Hotelling model, distance to nearest mine is one determinant of delivered coal price. Another determinant is distance to other nearby mines. As long as such distances are uncorrelated with distance to the nearest mine, the presence of other mines should not bias my results. As a test, I reconstruct my distance-based proxy as average distance to the nearest two and three mines in Columns (2) and (3), respectively.[24] Estimates are very similar to that of my benchmark suggesting that the influence of other nearby mines is not generating bias.

In Section 4.2, I discussed that reported generator characteristics are remarkably consistent across 1990-2012 EIA 860 forms. As a consequence, my constructed capacity dataset uses values from the latest year in which a generator is reported across the 1990-2012 EIA 860 forms to ensure that I observe all retired generators that were reported since 1990. To demonstrate that my results are not driven by this particular choice, in Column (4) I construct my capacity dataset using only the 2000 EIA 860 form while in Column (5) I use only the 2005 EIA 860 form. Results are similar to my benchmark estimate.

Other modeling choices Table 6 examines robustness to other modeling choices. For comparison, the benchmark result is reproduced in Column (1). Standard errors clustered at the county-level assumes that errors are not spatially correlated across counties. In Column (2), I reestimate my benchmark model with standard errors clustered at the state-year level to allow for arbitrary heteroskedasticity and spatial correlation across counties in the same state and year. Standard errors are very similar to my benchmark result. In Column (3), I examine what happens when I do not adjust coal and non-coal capacity by adding a unit value. I still detect statistically significant lagged effects three to five decade after the event though later effects, while mostly negative, are no longer precisely estimated. This may be due to the dramatically reduced sample size as zero values for non-coal capacity create indeterminate values in relative coal capacity. In Column (4), I consider a linear model of adjusted relative coal capacity, showing precisely estimated lagged effects that are similar in shape to that of the Poisson model. Finally, in Column (5) I estimate a linear model using raw unadjusted capacity values. Despite the much smaller sample, I still detect statistically significant lagged effects one, two, five, and six decades after obsolescence.

To provide a visual summary of how all these robustness checks compare with my benchmark estimate, Figure 5 plots my benchmark point estimates and confidence interval along with point estimates from the 11 different set of point estimates across Tables 4, 5, and 6.[25] Nearly all coefficients from these robustness tests fall within the uncertainty of my benchmark estimate.

6 Mechanisms

Section 5 finds robust evidence that transitory delivered coal price shocks have increasingly negative effects lasting through two generations of subsequent new relative coal capacity. Different mechanisms can generate such dynamics. Without further detail on the relevant mechanism, it is difficult to determine how such historical evidence can inform upon future energy transitions in the U.S. and elsewhere.

In the broadest sense, path dependence occurs when past relative fuel prices shift the relative marginal product of capacity in subsequent periods. Prior literature highlights two mechanisms that can generate

[24]The timing of the event is defined as before, when a shallow nearest mine is replaced by a deep mine.
[25]Models from Table 3 are excluded because they include a different set of controls.

this effect within the localized context examined here: increasing returns to scale and the accumulation of some fuel-specific stock factor such as productivity or human capital. To examine these competing channels, I first briefly summarize several pertinent physical features of power plants. I then introduce a model of power plant production that nests scale and productivity accumulation effects. This model informs a series of empirical tests designed to isolate the relevant mechanism. Finally, I test several other mechanisms that fall outside this theoretical framework.

6.1 Features of an electric power plant

A typical fossil fuel power plant converts primary fuel, such as coal, into electricity through three physical transformations. The fuel is burned in a boiler to generate heat, typically in the form of steam. Next, heat is converted to mechanical work through a steam turbine, which, in the final stage, is converted to electricity via a electricity generator. Three features of electricity production are particularly noteworthy.

Fixed input proportions Most boilers are designed for a specific "primary" fuel and input volume (Avallone, Baumeister and Sadegh (2006), p. 875). While alternative fuels may be used to ignite the boiler, sustained combustion of non-primary fuels or the primary fuel at different volumes can result in large efficiency losses (Avallone, Baumeister and Sadegh (2006), p.871). This implies that boiler capital, once built, is generally not substitutable with other inputs and thus production occurs with fixed input proportions. This "clay"-like nature of boilers has long been recognized in energy economics (Komiya, 1962; Joskow, 1985, 1987; Atkeson and Kehoe, 1999).

Local returns to scale Power plant capital is typically modular: boilers can provide steam to multiple turbines and generators. As new generators are added, larger boilers are built to serve both new and existing generators, providing efficiency gains to both old and new capital. This spillover effect, together with larger boilers typically being more thermally efficient, implies the presence of increasing returns to scale.[26] However, any increasing returns to scale is likely to be local as there are thermodynamic limits to the amount of heat that can be extracted from any unit of fuel. Nerlove (1963) and Christensen and Greene (1976) provide seminal early estimates of increasing returns to scale in the electricity sector.

Substitutability across fuel-specific electricity Electricity from different fuels often has distinct characteristics, making them gross but imperfect substitutes (Acemoglu et al., 2012). For example, coal-fired generators provide base load generation that can meet a steady level of demand. However, coal-fired boilers cannot be easily ramped up or down in response to short-run demand. In contrast, natural gas or diesel-fired generators can be adjusted rapidly in response to real-time demand. Similarly, renewable energy sources such as solar and wind, may generate less reliable electricity due to the intermittent availability of the renewable energy.

6.2 Theory: scale and productivity effects in electricity production

I consider a production function for a power plant with elements that correspond to these three features: vintaged fuel-specific electricity capacity with fixed factor proportions, increasing returns to scale for fuel-

[26] Formally, suppose electricity Y is produced using capital X, via the production function, $Y(X) = X$. If a doubling of capacity is achieved by a new boiler that services generators of all vintages, then $Y(2X) = 2bX$ where $b > 1$ is the efficiency improvement, which exhibits increasing returns to scale.

specific capacity, and imperfect substitutability between electricity from coal and other fuels. Each element is represented by a separate tier in the production function.

The time index t captures the increment between each vintage of electricity capacity. At the top tier, the power plant produces the final good, electricity, Y_t, using two intermediate goods, Y_{ct} and Y_{nt}, representing electricity from coal and all other fuels respectively.[27] Specifically, it takes the following Constant Elasticity of Substitution form:

$$Y_t = \left(Y_{ct}^{(\epsilon-1)/\epsilon} + Y_{nt}^{(\epsilon-1)/\epsilon}\right)^{\epsilon/(\epsilon-1)} \tag{2}$$

where ϵ is the elasticity of substitution between electricity produced by the two intermediate goods. Following the discussion in Section 6.1, I assume that electricity generated by different energy inputs are gross but imperfect substitutes, $\epsilon > 1$. The price of the final good is normalized to 1.[28]

Fuel-specific electricity is produced over the middle and lower tiers that separately capture scale and productivity effects. In the middle tier, fuel-specific electricity is generated by combining output across two vintages of generators in a Cobb-Douglas function with scale parameter ψ.[29] This scale parameter captures the efficiency gains noted in Section 6.1 from having larger boilers that service multiple vintages of generators. At the lower tier, generators of vintage t using fuel j produce output by combining capital, X_{jt}, and fuel, E_{jt}, in fixed proportions as represented by a Leontief function.[30] These proportions are captured by productivity terms A_{Xjt} and A_{Ejt} representing capital and fuel productivity, respectively. The intermediate good production function is:

$$Y_{jt} = (\min[A_{Xjt}X_{jt}], A_{Ejt}E_{jt})^{\alpha}(\min[A_{Xjt-1}\delta X_{jt-1}, A_{Ejt-1}E_{jt-1}])^{\alpha} \quad \text{for} \quad \text{j}\in\{\text{c,n}\} \tag{4}$$

where δ is capital depreciation, $\alpha \in \{0,1\}$ is the fuel-specific electricity elasticity of input so that $\psi = 2\alpha$.[31]

To explore how scale and productivity effects could generate path dependence for otherwise similar intermediate sectors, suppose capital and fuel productivities are the same across the two intermediate production functions in period $t-1$, $A_{Xct-1} = A_{Xnt-1}$ and $A_{Ect-1} = A_{Ent-1}$. First, observe that efficient allocation in the lower production structure imply $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ and $A_{Xjt-1}\delta X_{jt-1} = A_{Ejt-1}E_{jt-1}$ for each vintage t and fuel j. Next, the power plant's first order conditions in period t, rewritten in terms of relative current vintage capacity, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$, is (see Appendix C.1 for full derivation):

$$\widetilde{X}_t = \widetilde{w}_t^{\frac{\epsilon}{\varphi-1}} \widetilde{X}_{t-1}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \widetilde{A}_{Xt}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \tag{5}$$

where $\widetilde{w}_t$ is the relative input price index[32] and $\varphi = (1-\alpha)(1-\epsilon) < 0$, from earlier assumptions. Because capital depreciation is assumed to be the same in both intermediate sectors, it does now show up in Eq. 5. Eq. 5 provides two channels through which past relative input prices $\widetilde{w}_{t-1}$ can affect current-vintage relative coal capacity. First, applying Eq. 5 recursively shows that past relative coal prices affects past relative coal capacity, $\widetilde{X}_{t-1}$, the second term in Eq. 5. This is the scale effect. Second, while not explicitly

[27]For simplicity, this implies that electricity produced by all other inputs are perfect substitutes.

[28]Specifically,

$$\left[p_{ct}^{1-\epsilon} + p_{nt}^{1-\epsilon}\right]^{\frac{1}{1-\epsilon}} = 1 \tag{3}$$

[29]Assuming a CES function at this level introduces an extra substitution parameter that complicates the empirical implications but would not alter the main conclusion.

[30]This is also known as "clay-clay" production.

[31]This allows for diminishing marginal product under varying returns to scale. Otherwise, the relative input demand curve becomes upward sloping. Furthermore, the assumption that returns to scale is constant for coal and non-coal capacity is examined in Table 7.

[32]Specifically, $w_{jt} = \frac{A_{Xjt}}{A_{Ejt}} z_{jt} + r_t$ where z_{jt} is the primary energy input price and r_t is the rental price of capital.

modeled, any form of sector-biased endogenous technical change (i.e. fuel-biased technical change human capital accumulation) would generate stock accumulation in relative productivities. This implies that past relative input prices could alter current-vintage relative productivities. This is the productivity effect.

To empirically isolate which of these two effects are more relevant for estimates of path dependence in Section 5, I turn next to a nested series of empirical tests utilizing the tiered structure of this power plant production function. First, I conduct power plant-level regressions for plants that only burn coal to recover the scale parameter. Because such estimates may be biased in the presence of productivity effects, I then turn to generator-level regressions to test for productivity effects.

6.3 Testing for scale effects at the power plant level

To recover the scale parameter, I follow the cost function approach developed initially by Nerlove (1963) and implemented by Christensen and Greene (1976) to estimate reutrns to scale in the electricity sector. To remove the influence of the elasticity of substitution parameter, I restrict my sample to power plants m in county i and state s with generators that primarily burn coal. Using cost data from the Utility Data Institute (UDI) from 1981-1999,[33] cost minimization of Eq. 4 implies the following regression of non-fuel cost (see Appendix C.2 for full derivation):

$$\ln \overline{\text{non\_fuel\_cost}}_{mis} = \frac{1}{\psi} \ln \overline{Y}_{mis} + \overline{\boldsymbol{Z}}_{mis}\theta + \epsilon_{mis} \tag{6}$$

where the bar indicates time-averaged variables over 1981-1999. My parameter of interest is the scale parameter ψ. $\boldsymbol{Z}_{mis}$ is a vector of cross-sectional controls intended to absorb cross-sectional differences in input prices and productivity. They include observed power plant-level delivered coal price from UDI, state fixed effects and the latitude and longitude of the county centroid. In addition, I control for differences across transmission girds by including NERC region fixed effects. Standard errors are clustered at the county level.

Table 7 displays estimates of ψ. Column (1) includes plants in our sample of Midwest counties (see Figure A.7) which only burn coal. I estimate a statistically significant scale parameter of 1.8. Column (2) includes a modified sample containing plants in my sample Midwest counties that indicate coal as the primary fuel.[34] While these plants may have some fuel switching capability, I find that they exhibit similar returns to scale as coal-only power plants, with a statistically significant scale parameter of 2.1.

Potential simultaneity bias in Eq. 6 has been noted as early as Nerlove (1963). In particular, electricity prices for regulated electric utilities are often set to cover a power plant's average costs such that electricity output may be correlated with unobserved determinants of non-fuel costs. To address this concern, I use past approximated delivered coal prices as an instrument for current electricity output. Specifically, my instrument is the interaction between county distance to the obsolete mine and number of decades since the mine became obsolete. For identification to be valid, my distance-based proxy for past delivered coal prices must affect current non-fuel costs only through current output. Specifically, my first stage regression is:

$$\ln \overline{Y}_{mis} = \kappa_1 \ln dist_i^o * sinceEvent_i + \kappa_2 \ln dist_i^o + \kappa_3 sinceEvent_i + \overline{\boldsymbol{Z}}_{mis}\theta + \mu_{mis} \tag{7}$$

Eq. 7 estimates the event-time varying effects of distance to the obsolete mine and is the cross-sectional analog to my panel estimator in Eq. 1. In particular, κ_1 captures the event-time varying effect of distance to

[33] This limited time coverage prohibits panel regressions over the 20th century to recover the scale parameter. It also implies that the scale parameter I recover may only be relevant for recent decades.

[34] According to the UDI, federal forms used in compiling their data make no designation of primary or secondary fuels burned at a plant site. For multi-fuel plants, primary fuel is established by calculating the energy input of each fuel used and then assigning the primary fuel to that with the highest energy input.

the obsolete mine since it became obsolete and is analogous to the slope of the lagged effects shown in Figure 5. Columns (3) and (4) of Table 7 show IV estimates that are statistically significant, differ little depending on power plant sample, and are similar in magnitude to my OLS estimates. Furthermore, these IV estimates are robust to the potential presence of a weak instrument. For both IV estimates, the p-value and confidence interval from a conditional likelihood ratio test (Moreira, 2003) strongly rejects a null that the coefficient on electricity output in the structural equation is zero. Henceforth, my preferred scale parameter estimate is $\psi = 1.66$ from Column (3) of Table 7.

Finally, the production function presented in Section 6.2 assumes that coal and non-coal electricity exhibit similar returns to scale. To examine this assumption, I estimate Eq. 6 for power plants that primarily burn natural gas and oil in Columns (5) and (6) of Table 7. Because there are few Midwestern gas and oil-fired plants in the UDI data, I turn instead to all such power plants across the U.S. I find scale parameters for gas and oil-fired plants that are statistically indistinguishable from that of coal-fired plants.

6.4 Testing for productivity effects at the generator level

If past coal prices directly affect current coal-specific capital productivity, the exclusion restriction assumption for the IV estimator in Section 6.3 would fail and the scale parameter would not be identified. To detect the presence of productivity accumulation effects, I turn to generator-level regressions where productivity effects are most likely to be isolated. A standard engineering measure of generator productivity is thermal efficiency, the ratio of heat from electricity produced to heat from fuel consumed, and corresponds to $A_{Ect} = \frac{Y_{ct}}{E_{ct}}$ in the theory above. Using cross-sectional data of generator variables from EIA 860 forms averaged for the period 1990-1995,[35] I estimate the following regression of thermal efficiency for generator g, in power plant m, county i, and state s:

$$\ln \overline{\frac{Y}{E}}_{gmis} = \omega_1 \ln dist_i^o * sinceEvent_i + \omega_2 \ln dist_i^o + \omega_3 sinceEvent_i + \overline{\boldsymbol{Z}}_{gmis}\theta + \epsilon_{gmis} \tag{8}$$

where the set of controls $\boldsymbol{Z}_{mis}$ includes the distance-based proxy for contemporary delivered coal price, state and NERC region fixed effects, and the latitude and longitude of the county centroid. As with Eq. 7, ω_1 in Eq. 8 examines whether distance to the obsolete mine has an increasing effect on generator efficiency after obsolescence. If so, this would suggest that distance to the obsolete mine induces subsequent productivity improvements. Results are shown in Table 8 for generators within my benchmark sample of Midwest counties. Columns (1) and (2) include coal-burning generators in power plants that only burn coal. Columns (3) and (4) expands the sample to coal-burning generators in power plants that primarily burn coal. Columns (2) and (4) augments Eq. 8 by additionally including age and capacity of the generator. I do not find that distance to the obsolete mine has a statistically significant effect on generator thermal efficiency. The strongest determinant of thermal efficiency is the age of the generator which likely reflects aggregate technological change and not local productivity improvements driven by past price shocks.

It is perhaps unsurprising that path dependence from local shocks does not appear to be driven by local productivity accumulation. Standard determinants of technological change are typically not relevant at the county level. For example, directed technical change in capital productivity requires altering incentives in the research sector, which is rarely confined to within a county. Similarly, if productivity effects occur through the accumulation of coal-specific human capital, such effects are unlikely localized given that electric utilities can distribute human capital across power plants operating in multiple counties.

[35] generator heat rates, the inverse of thermal efficiency, was recorded in the EIA 860 forms only for the 1990-1995 period.

6.5 Testing other mechanisms

The theoretical framework presented in Section 6.2 incorporates two potential drivers of path dependence, scale and productivity accumulation effects, both operating within a power plant. However, because my treatment is at the county level, there could be other drivers beyond a power plant that may generate path dependence in relative coal capacity. This section examines several additional mechanisms.

Electricity is consumed by the manufacturing and residential sectors. If these sectors also exhibit scale effects then a negative historic coal price shock may lead to sectoral expansion. Indeed, previous literature has detected long-term effects of historical access to hydropower electricity on later local manufacturing sector employment (Kline and Moretti, 2014) and population density (Severnini, 2014). Path dependence in relative coal capacity may result if coal-fired electricity is preferred in meeting this increased demand, possibly because of its advantage in providing base load electricity.

The first evidence that demand-side effects alone cannot explain for path dependence comes from Table 2 which detects lagged effects on not only coal capacity, but on relative coal capacity which normalizes for overall local electricity demand. To provide additional evidence, Table 9 shows the effect of distance to the obsolete mine since obsolescence on various county characteristics that partially capture electricity demand in the residential and manufacturing sectors. Specifically, I separately examine effects on total population, urban population, manufacturing establishments, and manufacturing employment in the eight decades after obsolescence, as indicated across Columns (1)-(8). To be consistent with the pre-trend regressions in Column (3) of Table 1, outcomes are in log first differences and all models include state fixed effects. While there are a few statistically significant coefficients, none of the county characteristics exhibit a systematic pattern of lagged effects like that found for relative coal capacity.

Coal-specific capital is not exclusive to a power plant. In order for coal to arrive at a power plant, rail and highway networks are needed as complementary capital. Increasing returns along with sunk costs in coal transport infrastructure can similarly generate path dependent dynamics in relative coal capacity.[36] Unfortunately, unlike the county outcomes examined in Table 9, county-level data over the 20th century for rail and roadway density is not available. Instead, I use measures of rail and highway density (miles per square mile) in 2010 in a county-level cross-sectional regression similar to the generator-level specification estimated in Eq. 8. Results are shown in Table 10. I find that neither forms of transport infrastructure responds to distance to the obsolete mine.

Finally, Joskow (1987) finds that power plants procure coal using long-duration contracts. The presence of long-term contracts may induce plants to continue purchasing historic sources of coal even as contemporaneous circumstances change. There are two reasons why such procurement practices would not fully explain my estimates of path dependence. First, Joskow (1987) showed that the average coal contract length in 1979 lasted 12.8 years with a standard deviation of 10.4 years. Using EIA 423 data over 1983-1997, Jha (2015) finds this duration has decreased recently to 4.4 years with a standard deviation of 5.9 years. Coal contracts of such duration may induce path dependence in relative coal capacity within a decade but is unlikely to generate lagged effects over multiple decades. Furthermore, even if coal contracts were of longer duration, it is unclear why such contracts would cause lagged effects to become more negative over time.

[36] Sunk costs alone, however, would not generate path dependence in the long-run as the capital would eventually depreciate. See discussion in Bleakley and Lin (2012).

7 Structural interpretation

Detecting increasing returns to scale as the mechanism behind path dependence in relative coal capacity allows reduced-form estimates in Section 5 to be interpreted structurally. First, I formally define path dependence strength as a function of two parameters: returns to scale, ψ, and the elasticity of substitution between coal and non-coal electricity, ϵ. Second, using this definition, I recover the elasticity of substitution as implied by estimates of path dependence in Section 5 and the scale parameter in Section 6.3. Finally, to explore implications for national U.S. emissions, I apply these structural parameters in simulations of future electricity CO_2 emissions following relative coal price shocks of varying magnitude and duration.

7.1 Strength of path dependence: formal definition

In the absence of productivity accumulation effects, past relative input prices affect relative capacity only through past relative capacity.[37] Applying a natural log transformation to Eq. 5 and rewriting recursively, current vintage relative coal capacity is:

$$\begin{aligned}
\ln \widetilde{X}_t &= \frac{\epsilon}{(\varphi-1)} \ln \widetilde{w}_t + \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} \ln \widetilde{w}_{t-1} + \frac{\alpha^2(1-\epsilon)^2\epsilon}{(\varphi-1)^3} \ln \widetilde{w}_{t-2} + ... \\
&\quad + \frac{\alpha(1-\epsilon)}{(\varphi-1)} \ln \widetilde{A}_{Xt} + \frac{\alpha^2(1-\epsilon)^2}{(\varphi-1)^2} \ln \widetilde{A}_{Xt-1} + \frac{\alpha^3(1-\epsilon)^3}{(\varphi-1)^3} \ln \widetilde{A}_{Xt-2} + ... \\
&= \sum_{s=0}^{\infty} \frac{\epsilon}{(\varphi-1)} \left[\frac{\alpha(1-\epsilon)}{(\varphi-1)}\right]^s \ln \widetilde{w}_{t-s} + \sum_{s=0}^{\infty} \left[\frac{\alpha(1-\epsilon)}{(\varphi-1)}\right]^{s+1} \ln \widetilde{A}_{Xt-s} \qquad (9)
\end{aligned}$$

where s is the lagged time index. Higher relative coal prices lower contemporaneous relative coal capacity, $\frac{\partial \ln \widetilde{X}_t}{\partial \ln \widetilde{w}_t} < 0$. Higher past relative coal prices also lower current relative coal capacity, $\frac{\partial \ln \widetilde{X}_t}{\partial \ln \widetilde{w}_{t-s}} < 0$. The strength of path dependence depends on the relative magnitude between contemporaneous and lagged effects. Formally, I define:

PROPOSITION 1 ***Weak path dependence:*** *The effect of past relative coal prices weakens over time,* $\frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_{t-1})} - \frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_t)} = \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} - \frac{\epsilon}{(\varphi-1)} > 0$*, or when* $\psi < \frac{-\epsilon}{1-\epsilon}$.

PROPOSITION 2 ***Strong path dependence:*** *The effect of past relative coal prices strengthens over time,* $\frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_{t-1})} - \frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_t)} = \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} - \frac{\epsilon}{(\varphi-1)} < 0$*, or when* $\psi > \frac{-\epsilon}{1-\epsilon}$.

Strong path dependence occurs whenever an increase in relative coal prices triggers a downward shift in the marginal product curve of relative coal capacity that continues in each subsequent period. Two countervailing forces are at work. Consider first only the role of the scale parameter ψ on coal-fired capacity. When $\psi > 1$, the cross partial derivative of Y_{ct} exceeds the second partial derivative of Y_{ct} with respect to current coal capacity such that an increase in relative coal prices has a stronger effect on subsequent coal capacity than on current coal capacity.[38] However, in the presence of multiple intermediate sectors, $\psi > 1$ alone does not

[37] Note that there may still be path dependence due directly from exogenous productivity shocks. However, evidence from Section 6 suggests it is unlikely that past relative coal prices are affecting current relative coal capacity through productivity effects.

[38] This results from a straightforward application of Euler's theorem. Formally, the cross partial derivative of a function $Y(X_t, X_{t-1})$ of homogeneous degree ψ can be written as:

dictate the strength of path dependence in relative coal capacity. A countervailing force comes from the imperfect substitutability between coal and non-coal electricity with the degree of substitutability, $\epsilon > 1$, determining how much relative electricity prices favor the less abundant fuel-specific electricity. An increase in the relative coal price induces a contemporaneous decrease in relative coal capacity. However, because vintaged-capital is durable, there is now an excess supply of non-coal capacity in the subsequent period after the shock. The resulting higher price for coal-fired electricity induces a relative increase in coal capacity such that, over time, relative coal capacity converges back to the pre-shock level. Thus, strong path dependence can only be achieved through a combination of strong increasing returns to scale and a high degree of substitutability between coal and non-coal electricity.[39] In the case where coal and non-coal electricity are weak substitutes (i.e. a low ϵ or high $\frac{-\epsilon}{1-\epsilon}$), ψ must be sufficiently large such that the forces of increasing returns offset that of imperfect substitutability. Put in another way, strong path dependence in electricity sector transitions occur when there are strong increasing returns to scale in production and when electricity produced by different primary fuels exhibits similar properties.

7.2 Recovering the elasticity of substitution

The elasticity of substitution between coal and non-coal electricity, ϵ, is a key parameter found across a broad class of multi-sector structural change models (Baumol, 1967; Ngai and Pissarides, 2007; Acemoglu, 2002; Acemoglu and Guerrieri, 2008; Acemoglu et al., 2012; Lemoine, 2016). A recovered value for this parameter can be used to inform other models of energy transition featuring other drivers of structural change. Specifically, this elasticity is recovered by mapping the structural expression in Eq. 9 to the reduced-form regression specification in Eq. 1.

Each lagged effect in Eq. 9 is expressed in terms of vintages. The reduced-form analog is the lagged effect of distance to the obsolete mine on the next generation of gross new relative capacity, as shown in Column (6) of Table 2.[40] To account for uncertain capacity lifespan (see Figure A.15), I weight coefficients $\widehat{\beta^o_{\tau=1}}$ through $\widehat{\beta^o_{\tau=5}}$ by the likelihood of each lifespan to obtain a lifespan weighted lagged effect on new relative capacity, $\frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} - \frac{\epsilon}{(\varphi-1)} = -2.84$.[41] Together with the estimated scale parameter, $\psi = 1.66$, from Section 6.3, this implies $\epsilon = 4.9$.[42] As an example of how this elasticity value can inform other structural change models, in Acemoglu et al. (2012)'s model of optimal climate policy under directed technical change, $\epsilon = 4.9$ falls within the parameter space for which a temporary policy intervention is sufficient to avoid climate change disaster.

$$\frac{\partial^2 Y_{ct}}{\partial X_t \partial X_{t-1}} = \left(\frac{\psi - 1}{X_{t-1}}\right)\frac{\partial Y_{ct}}{\partial X_t} - \left(\frac{X_t}{X_{t-1}}\right)\frac{\partial^2 Y_{ct}}{\partial^2 X_t^2}$$

Setting $X_t = X_{t-1}$ so that one can compare the effects of lagged capacity against current capacity, it is evident that $\frac{\partial^2 Y_{ct}}{\partial X_t \partial X_{t-1}} > -\frac{\partial^2 Y_{ct}}{\partial^2 X_t^2}$ only in the presence of increasing returns to scale, or when $\psi > 1$.

[39] Notice the similarities here with the "market" and "price" effects found in standard models of directed technical change (Acemoglu, 2002; Acemoglu et al., 2012).

[40] I consider only the effect on the first, and not second, subsequent generation of gross new relative coal capacity because the first generation effect is estimated from a larger set of counties that experienced treatment throughout the 20th century.

[41] Specifically, the next generation new relative capacity effect is $.12(-2.1) + .10(-2.6) + .33(-4.2) + .24(-2.7) + .21(-1.4) = -2.84$. The weights come from Figure A.15. The standard error, which accounts for uncertainty in $\widehat{\beta^o_\tau}$, is 1.17.

[42] To solve for ϵ analytically, one can rewrite $\frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} - \frac{\epsilon}{(\varphi-1)} = -2.84$ in the following quadratic form:

$$\epsilon^2[2(1+2.84)\alpha - (1+2.84) - 2.84\alpha^2] + \epsilon[-2(1+2.84)\alpha + 2 * 2.84\alpha^2] - 2.84\alpha^2 = 0$$

where the positive root of the above equation is the implied ϵ.

7.3 Simulating U.S. emissions pathways following temporary price shocks

Evidence of strong path dependence implies that it is possible for a temporary price shock to induce permanent fuel switching. The magnitude and/or duration of the required shock, however, depends on baseline relative coal prices in the absence of the shock. If baseline coal prices are low, a large and/or long lasting price shock is needed to prevent the forces of path dependence from once again favoring coal after the shock dissipates.

To analyze how a temporary price shock could overcome baseline relative coal prices at the national level, I use my model and estimated structural parameters to simulate future U.S. electricity sector CO_2 emissions following shocks of varying magnitude and duration. To ground my simulations in recent developments, I consider relative coal price shocks based on recent advances in hydraulic fracturing which have increased natural gas supplies in the U.S. and elsewhere. Figure A.16 plots the ratio of log coal to log natural gas price for U.S. industrial consumers from 1970-2011. As a consequence of hydraulic fracturing, relative coal prices since 2009 have increased by 98% relative to the historic trend.

My national-level simulations employ the following simplifying assumptions. First, each time step is now indexed by decade because new vintaged capacity is built more frequently at the national level. Second, coal-fired capacity can only be substituted for natural-gas fired capacity.[43] Third, to avoid forecasting trends in productivity and overall electricity demand, I assume that future relative fuel-specific productivity and total coal and natural gas capacity are held at 2000-2009 levels. Fourth, I assume that in the absence of the fracturing-induced relative price shock, baseline relative coal prices and capital depreciation rates are held at 2000-2009 levels. Finally, I assume that the scale and elasticity of substitution parameters are constant over the simulation period. Because of these assumptions, it is important to note that these simulations are meant not as forecasts of U.S. emissions but rather as an exercise in comparing how the dynamics of path dependence stack up against the role of baseline U.S. coal prices. Appendix D details the simulation procedure.

Figure 7 shows emissions pathways from varying two features of the temporary shock. Simulations down row panels employ shocks that are 1, 1.5, and 3 times larger than recent relative coal prices due to hydraulic fracturing. Simulations across column panels allow these shocks to last 1, 2, and 3 decades. The baseline emissions pathway in the absence of the shock is displayed as a dashed gray line. The emissions pathway and share of new coal capacity under mean parameter values are shown as bold blue and red lines, respectively. To incorporate estimated uncertainty in my structural parameters, ψ and ϵ, I also plot the emissions pathway and share of new coal capacity after drawing values from each parameter's empirical distribution. These are displayed as thin blue and red lines, respectively.

While any temporary shock lowers emission levels relative to the baseline case, achieving a long-term emissions decline is highly unlikely even if recent prices from hydraulic fracturing last multiple decades (top row). If the shock were 1.5 times larger than recent prices, it still must last 3 decades for a better than 50% chance of declining long-term emissions (middle row). A 3 times larger shock would reduce the required duration to 2 decades (bottom row). In none of the simulations considered do emissions reach zero in the long run, but rather converge asymptotically to a steady-state level where all electricity is produced by natural gas. This is because natural gas still contains carbon, albeit less than coal.

[43]In 2009, coal and natural gas constituted 92% of total U.S. fossil fuel-fired electricity capacity.

8 Conclusion

This paper estimates the strength of path dependence in the electricity sector for the U.S. Midwest over the 20th century. Exploiting variation in county delivered coal prices driven by the introduction of mechanized mining, I find that transitory price shocks have effects that increase in magnitude over two subsequent generations of relative coal-fired electricity capacity. Additional evidence finds increasing returns to scale in electricity production as the likely mechanism causing path dependence. Interpreted through a model of electricity production, my reduced-form estimates map onto an elasticity of substitution between coal and non-coal electricity value of 5.

This historical evidence of path dependence in the electricity sector is particularly timely given the possible energy transition currently underway in the U.S. The recent spike in relative coal prices due to technology breakthroughs in natural gas extraction has led to retirements of existing coal-fired capacity and proposed construction of new natural gas-fired capacity. Simulations of future emissions using estimated structural parameters show that a permanent decline in U.S. electricity sector emissions would require shocks of larger magnitude and/or longer duration than recently observed under hydraulic fracturing.

Under technology-specific policies, path dependence could actually increase the ultimate cost of emissions abatement. This is because path dependence cuts both ways by amplifying the long-term consequences of picking either the right or wrong technology. Suppose natural gas prices are low and lasting enough to induce an endogenous transition from coal to natural gas-fired electricity. If climate damages turn out to be so large that optimal mitigation requires cleaner fuel than natural gas, the subsequent path dependence in natural gas would increase the cost of switching to a cleaner fuel than if the detour into natural gas had been avoided. The presence of path dependence therefore provides an added argument in favor of Pigouvian interventions that directly price an externality over technology-specific policies.

References

Acemoglu, Daron. 2002. "Directed Technical Change." *The Review of Economic Studies*, 69(4): 781–809.

Acemoglu, Daron, and Veronica Guerrieri. 2008. "Capital Deepening and Nonbalanced Economic Growth." *Journal of Political Economy*, 116(3): 467–498.

Acemoglu, Daron., Philippe. Aghion, Leonardo Bursztyn, and David Hemous. 2012. "The Environment and Directed Technical Change." *American Economic Review*, 102(1): 131–66.

Acemoglu, Daron, Ufuk Akcigit, Douglas Hanley, and William Kerr. 2016. "Transition to Clean Technology." *Journal of Political Economy*, 124(1): 52–104.

Aghion, Philippe, and Peter Howitt. 1992. "A Model of Growth Through Creative Destruction." *Econometrica*, 60(2): pp. 323–351.

Aghion, Philippe, Antoine Dechezlepretre, David Hemous, Ralf Martin, and John Van Reenen. 2016. "Carbon Taxes, Path Dependency, and Directed Technical Change: Evidence from the Auto Industry." *Journal of Political Economy*, 124(1): 1–51.

Arthur, W. Brian. 1994. *Increasing Returns and Path Dependence in the Economy.* University of Michigan Press.

Atkeson, Andrew, and Patrick J. Kehoe. 1999. "Models of Energy Use: Putty-Putty versus Putty-Clay." *American Economic Review*, 89(4): 1028–1043.

Avallone, Eugene A, Theodore Baumeister, and Ali Sadegh. 2006. *Marks' Standard Handbook For Mechanical Engineers (Standard Handbook for Mechanical Engineers).* Mcgraw-Hill Professional.

Barreca, Alan, Karen Clay, and Joel Tarr. 2014. "Coal, Smoke, and Death: Bituminous Coal and American Home Heating." National Bureau of Economic Research Working Paper 19881.

Baumol, William J. 1967. "Macroeconomics of unbalanced growth: the anatomy of urban crisis." *American Economic Review*, 415–426.

Beach, Brian, and W. Walker Hanlon. 2016. "Coal Smoke and Mortality in an Early Industrial Economy." mimeo.

Bleakley, Hoyt, and Jeffrey Lin. 2012. "Portage and Path Dependence." *The Quarterly Journal of Economics*, 127(2): 587–644.

Boden, T.A., G. Marland, and R.J. Andres. 2013. "Global, Regional, and National Fossil-Fuel CO2 Emissions." Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, Tenn., U.S.A. Working Paper.

Campbell, Marius R. 1908. "Coal Fields of the United States." U. S. Geological Survey Map.

C. d'Aspremont, J. Jaskold Gabszewicz, J.-F. Thisse. 1979. "On Hotelling's "Stability in Competition"." *Econometrica*, 47(5): 1145–1150.

Chay, Kenneth Y., and Michael Greenstone. 2003. "The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession." *The Quarterly Journal of Economics*, 118(3): 1121–1167.

Chay, Kenneth Y., and Michael Greenstone. 2005. "Does Air Quality Matter? Evidence from the Housing Market." *Journal of Political Economy*, 113(2): pp. 376–424.

Christensen, Laurits R, and William H Greene. 1976. "Economies of scale in US electric power generation." *The Journal of Political Economy*, 655–676.

Clay, Karen, Joshua Lewis, and Edson Severnini. 2016. "Canary in a Coal Mine: Infant Mortality, Property Values, and Tradeoffs Associated with Mid-20th Century Air Pollution." National Bureau of Economic Research Working Paper 22155.

Coase, R. H. 1960. "The Problem of Social Cost." *Journal of Law and Economics*, 3: pp. 1–44.

David, Paul A. 1985. "Clio and the Economics of QWERTY." *The American Economic Review*, 75(2): pp. 332–337.

Dell, Melissa. 2012. "Path Dependence in Development: Evidence from the Mexican Revolution." mimeo.

Devine, Warren D., Jr. 1983. "From Shafts to Wires: Historical Perspective on Electrification." *The Journal of Economic History*, 43(2).

East, J.A. 2012. "Coal fields of the conterminous United States, National Coal Resource Assessment." U.S. Geological Survey 2012-1205.

Energy Information Administration. 1980. "Inventory of Power Plants in the United States."

Energy Information Administration. 2012. "Annual Energy Review, 1982-2011."

Energy Information Administration. 2013. "Annual Energy Outlook."

Energy Information Administration. 2015. "Monthly Energy Review."

Federal Power Commission. 1941. "Plant and Ownership List to Accompany Map 'Principal Electric Utility Generating Stations and Transmission Lines'." Government Printing Office, Washington, DC.

Federal Power Commission. 1945. "Plant and Ownership List to Accompany Map 'Principal Electric Utility Generating Stations and Transmission Lines'." Government Printing Office, Washington, DC.

Federal Power Commission. 1951. "Plant and Ownership List to Accompany Map 'Principal Electric Utility Generating Stations and Transmission Lines'." Government Printing Office, Washington, DC.

Federal Power Commission. 1954. "Plant and Ownership List to Accompany Map 'Principal Electric Utility Generating Stations and Transmission Lines'." Government Printing Office, Washington, DC.

Federal Power Commission. 1963. "Principal electric power facilities in the United States (map)." Government Printing Office, Washington, DC.

Feyrer, James. 2009. "Trade and Income – Exploiting Time Series in Geography." National Bureau of Economic Research Working Paper 14910.

Fisher, Cassius A. 1910. "Depth and Minimum Thickness of Beds as Limiting Factors in Valuation." United States Geological Survey Bulletin 424.

Gaudet, Grard, Michel Moreaux, and Stephen W. Salant. 2001. "Intertemporal Depletion of Resource Sites by Spatially Distributed Users." *The American Economic Review*, 91(4): pp. 1149–1159.

Glaeser, Edward L, Sari Pekkala Kerr, and William R Kerr. 2015. "Entrepreneurship and urban growth: An empirical assessment with historical mines." *Review of Economics and Statistics*, 97(2): 498–520.

Haines, Michael R, and Inter-university Consortium for Political and Social Research. 2010. "Historical, Demographic, Economic, and Social Data: The United States, 1790-2002." *Inter-university Consortium for Political and Social Research (ICPSR) [distributor].*

Hanlon, W. Walker. 2016. "Coal Smoke and the Costs of the Industrial Revolution." mimeo.

Herfindahl, Orris C. 1967. "Depletion and economic theory." *Extractive resources and taxation*, 63–90.

Hornbeck, Richard. 2012. "The Enduring Impact of the American Dust Bowl: Short- and Long-Run Adjustments to Environmental Catastrophe." *American Economic Review*, 102(4): 1477–1507.

Hornbeck, Richard, and Pinar Keskin. 2014. "The Historically Evolving Impact of the Ogallala Aquifer: Agricultural Adaptation to Groundwater and Drought." *American Economic Journal: Applied Economics*, 6(1): 190–219.

Hornbeck, Richard, and Suresh Naidu. 2014. "When the Levee Breaks: Black Migration and Economic Development in the American South." *American Economic Review*, 104(3): 963–90.

Hotelling, Harold. 1929. "Stability in Competition." *The Economic Journal*, 39(153): pp. 41–57.

Jha, Akshaya. 2015. "Regulatory Induced Risk Aversion: Coal Procurement at U.S Power Plants." mimeo.

Joskow, Paul L. 1985. "Vertical Integration and Long-Term Contracts: The Case of Coal-Burning Electric Generating Plants." *Journal of Law, Economics, & Organization*, 1(1): 33–80.

Joskow, Paul L. 1987. "Contract duration and relationship-specific investments: Empirical evidence from coal markets." *The American Economic Review*, 168–185.

Kline, Patrick, and Enrico Moretti. 2014. "Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority." *The Quarterly Journal of Economics*, 129(1): 275–331.

Knittel, Christopher R., Konstantinos Metaxoglou, and Andre Trindade. 2015. "Natural Gas Prices and Coal Displacement: Evidence from Electricity Markets." National Bureau of Economic Research Working Paper 21627.

Kolstad, Charles D. 1994. "Hotelling Rents in Hotelling Space: Product Differentiation in Exhaustible Resource Markets." *Journal of Environmental Economics and Management*, 26(2): 163 – 180.

Komiya, Ryutaro. 1962. "Technological Progress and the Production Function in the United States Steam Power Industry." *The Review of Economics and Statistics*, 44(2): 156–166.

Lemoine, Derek. 2016. "Innovation-Led Transitions in Energy Supply." mimeo.

Lewis, Joshua. 2014. "Short-run and Long-run Effects of Household Electrification." mimeo.

McNerney, James, J. Doyne Farmer, and Jessika E. Trancik. 2011. "Historical costs of coal-fired electricity and implications for the future." *Energy Policy*, 39(6): 3042 – 3054.

Moreira, Marcelo J. 2003. "A conditional likelihood ratio test for structural models." *Econometrica*, 71(4): 1027–1048.

Nerlove, Marc. 1963. "Returns to Scale in Electricity Supply." In *Measurement in Economics - Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld.* , ed. Carl F. Christ. Stanford, CA:Stanford Univ. Press.

Ngai, L. Rachel, and Christopher A. Pissarides. 2007. "Structural Change in a Multisector Model of Growth." *American Economic Review*, 97(1): 429–443.

Nunn, Nathan, and Nancy Qian. 2011. "The Potato's Contribution to Population and Urbanization: Evidence From A Historical Experiment." *The Quarterly Journal of Economics*.

Redding, Stephen J, Daniel M Sturm, and Nikolaus Wolf. 2011. "History and Industry Location: Evidence from German Airports." *Review of Economics and Statistics*, 93(3): 814–831.

Schmookler, Jacob. 1966. *Invention and Economic Growth.* Harvard University Press.

Schumpeter, J. 1942. *Capitalism, Socialism and Democracy.* Harper.

Severnini, Edson. 2014. "The Power of Hydroelectric Dams: Agglomeration spillovers." IZA Discussion Paper.

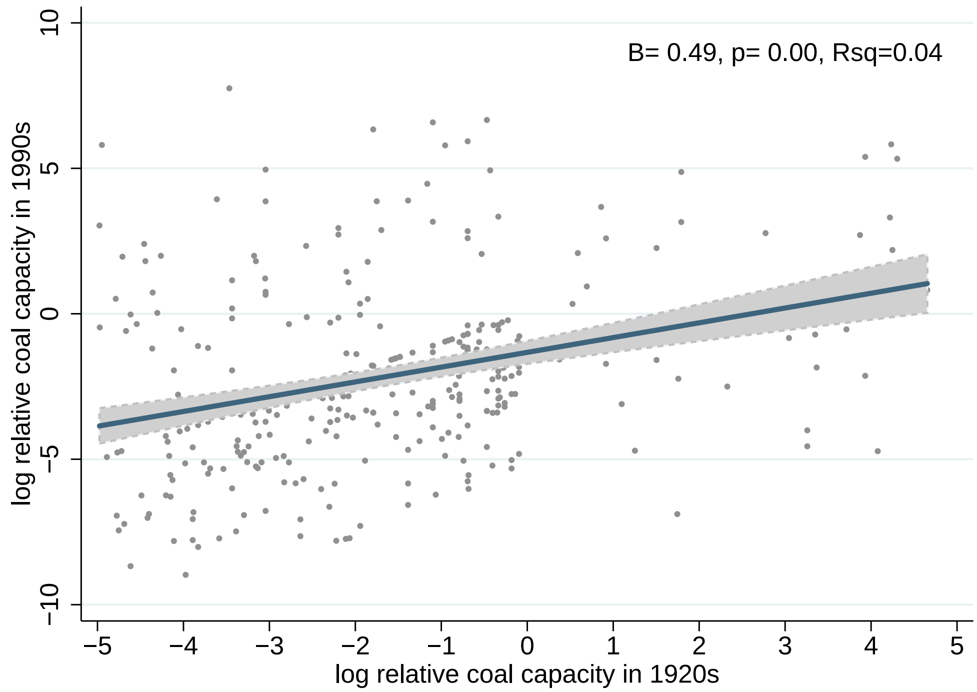
Speight, J. G. 1994. *The Chemistry and Technology of Coal, 2nd edition.* Marcel Dekker, Inc.

U.S. Census Bureau. 1975. "Historical Statistics of the United States, Colonial Times to 1970, Bicentennial Edition." U.S. Department of Commerce U.S. Geological Survey.

World Bank. 2014. "Data retrieved March, 2014, from World Development Indicators Online (WDI) database."

Figures

Figure 1: County relative coal capacity in the 1990s versus the 1920s

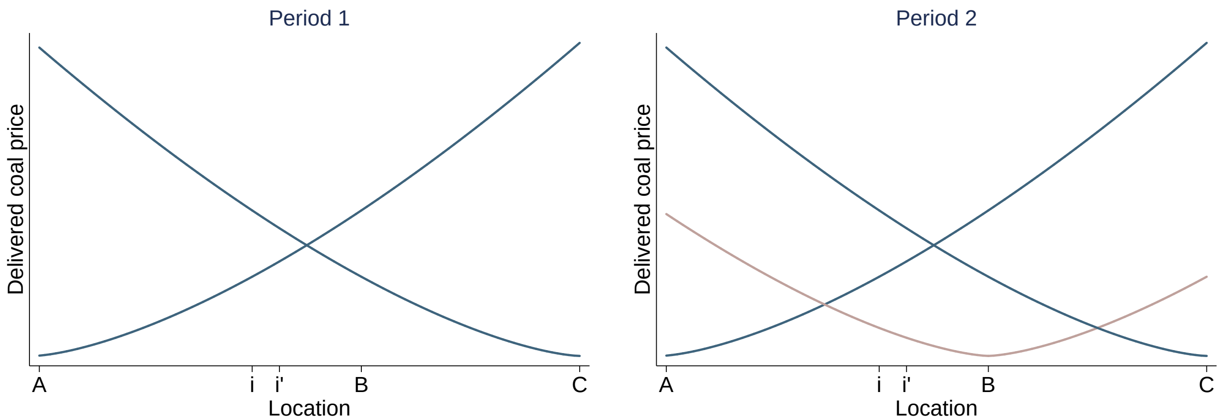


Notes: Scatter shows county-level log relative coal capacity in the 1990s against log relative coal capacity in the 1920s. Data from EIA 860 forms (see Section 4.2).

Figure 2: County relative coal capacity over 20th century for counties close to and far from Illinois Basin

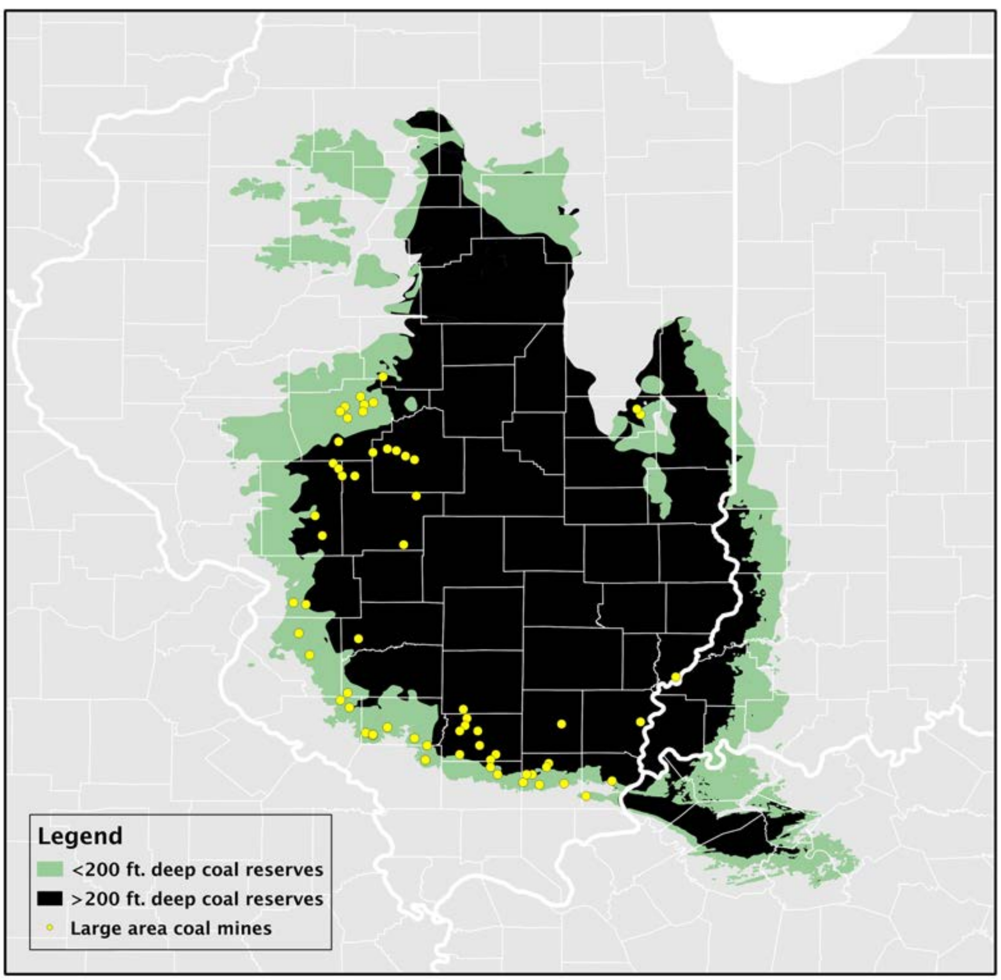
Notes: Average log relative coal capacity over time for counties less than 50 miles (solid blue) and between 200 and 250 miles (dashed gray) from the Illinois Coal Basin. Illinois Coal Basin data from East (2012) (see Section 4.1). Capacity data from EIA 860 forms (see Section 4.2).

Figure 3: Hotelling location model schematic



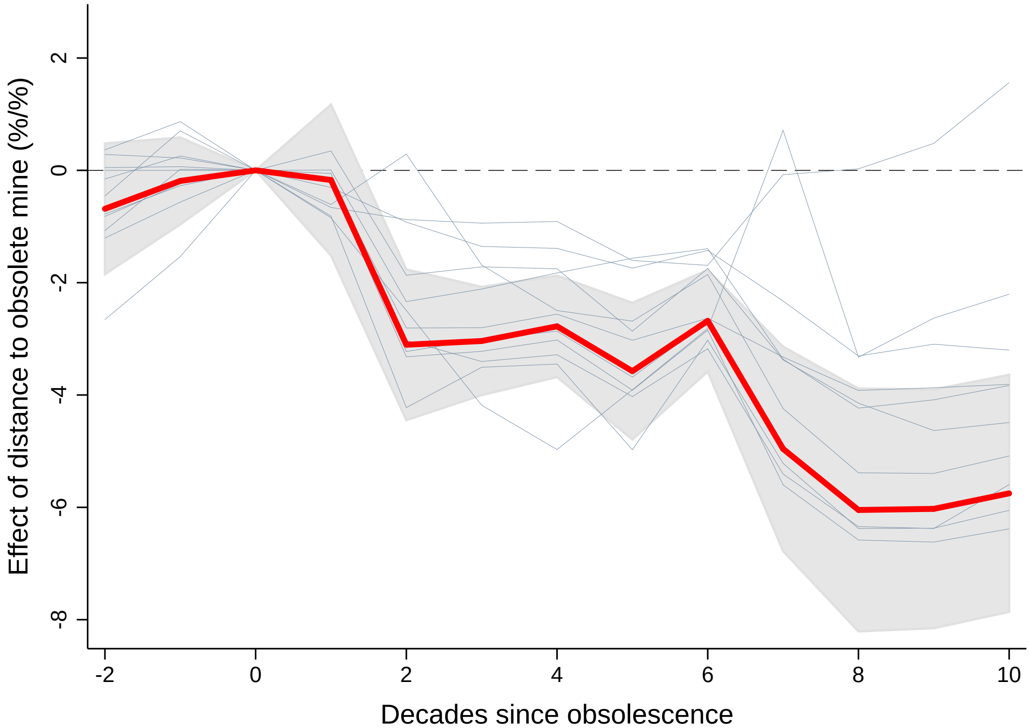
Notes: Delivered prices under a 1-dimensional Hotelling location model. Lines show delivered coal prices from mines in locations A, B, and C. See Section 3.2 for discussion.

Figure 4: Map of Illinois coal basin by resource depth and mine location



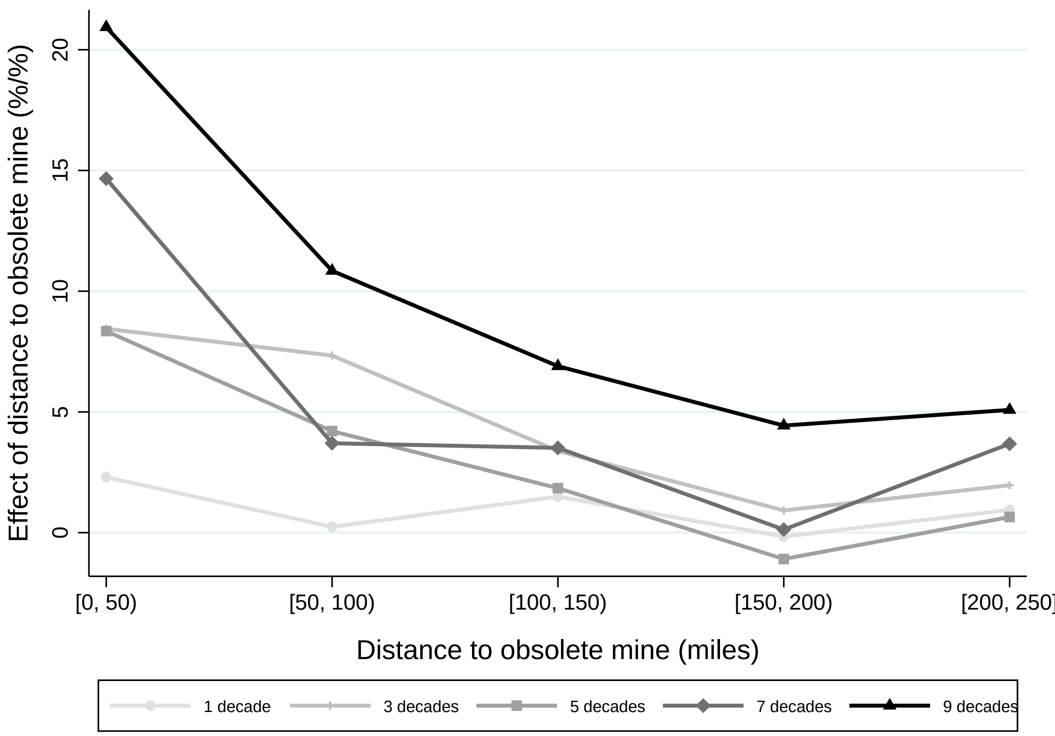
Notes: Lighter shaded area indicates the location of coal resources less than 200 feet deep. Darker shaded area indicates the location of coal resources more than 200 feet deep. Yellow points show location of large coal mines that operated at any point after 1880. County and state boundaries shown. Source: East (2012).

Figure 5: Estimates of path dependence on relative coal capacity



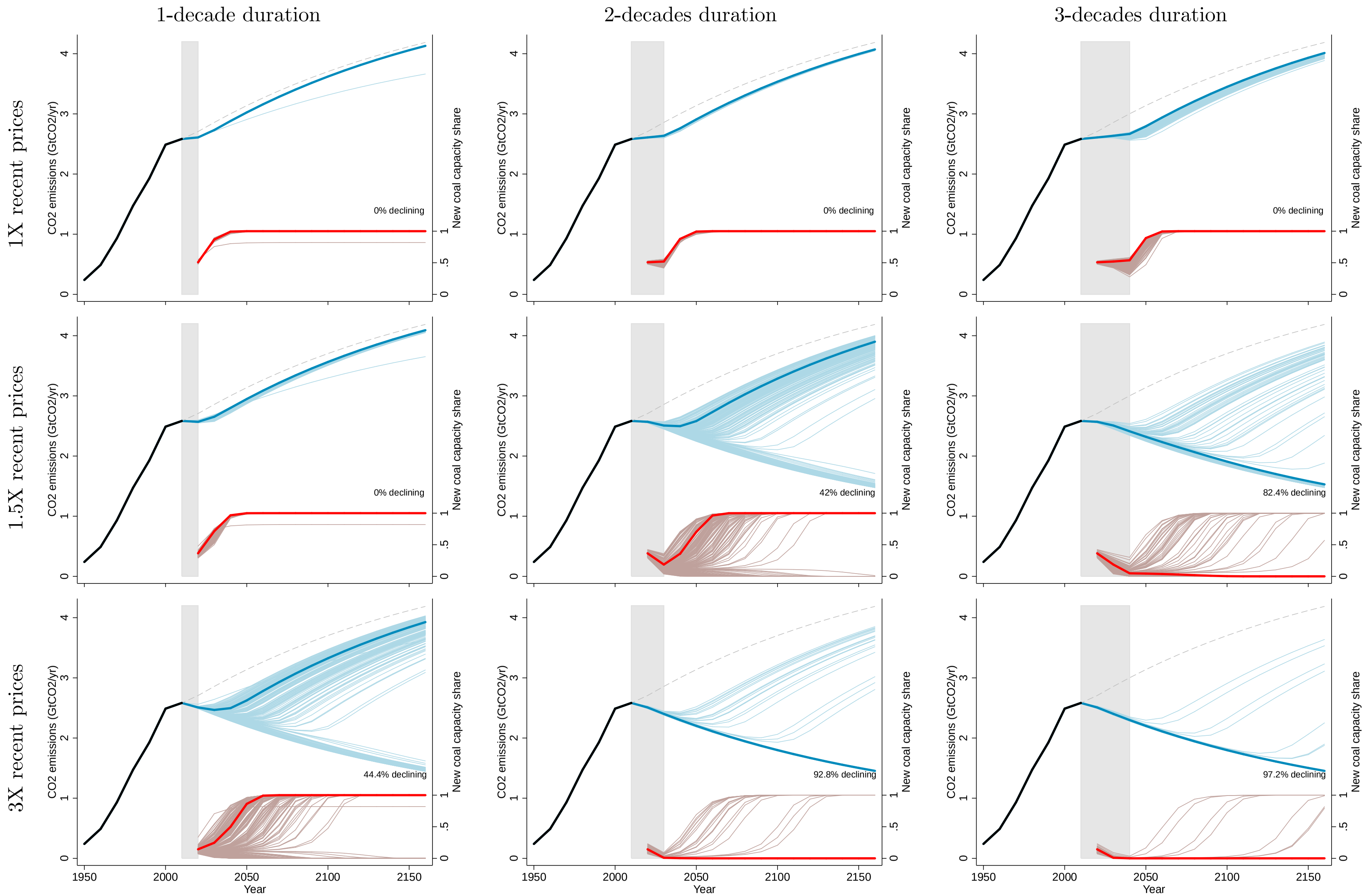
Notes: Solid red line and shaded gray area shows coefficients β^o_τ and 90% confidence intervals corresponding to the benchmark model from Column (5) in Table 2. Each thin blue line shows coefficients from the 11 robustness checks in Tables 4, 5, and 6.

Figure 6: Testing for nonlinearity



Notes: Plot examines nonlinearity in lagged effects of distance to obsolete coal mine. Coefficients shown one, three, five, and seven decade lagged effects broken into 50-mile wide dummies.

Figure 7: Simulating future U.S. CO_2 emissions following temporary relative coal price shocks



Notes: Simulations of U.S. CO_2 electricity sector emissions from coal and natural gas-fired capacity following temporary relative coal price shocks. Top, middle, and bottom panels use price shocks 1, 1.5 and 2 times that of recent relative coal prices. Left, middle, and right panels use price shocks that last 1, 2, and 3 decades (gray area). Solid black line shows historic emissions. Dashed gray line shows future business-as-usual emissions without price shock. Solid blue (red) line shows projected emissions (new coal capacity share) using point estimates of scale and elasticity of substitution parameters. Thin blue (red) lines shows projected emissions (new coal capacity share) using 250 Monte Carlo draws from the estimated distributions of scale and elasticity of substitution parameters. Percentage of draws with declining emissions pathways shown.

Tables

Table 1: Examining pre-trends in county characteristics

	(1)	(2)	(3)
Outcome	Levels	Changes	
Log population (1890-1990)	-0.13	-0.014	-0.016
	[0.086]	[0.0090]	[0.013]
Number of counties	261	261	261
Log urban population (1890-1980)	-0.085	-0.013	0.019
	[0.17]	[0.017]	[0.022]
Number of counties	183	171	171
log mgf establishments (1890-1990)	-0.20**	-0.052	0.046
	[0.098]	[0.056]	[0.064]
Number of counties	260	260	260
log mgf employment (1890-1990)	-0.16	0.11**	0.00096
	[0.20]	[0.052]	[0.068]
Number of counties	258	255	255
log mfg capital (1890-1900)	0.071	0.17**	0.093
	[0.24]	[0.085]	[0.15]
Number of counties	106	105	105
log mfg output (1890-1940)	-0.44	0.17**	0.044
	[0.35]	[0.075]	[0.16]
Number of counties	114	113	113
State fixed effects	No	No	Yes

Notes: Each coefficient from a separate regression showing effect of log distance to eventual obsolete mine on county-level characteristics when mine was active price setter. Time coverage for county characteristics indicated below variable name. Counties from benchmark sample (see Figure A.7). Outcomes in Column (1) in logs. Outcomes in Columns (2)-(3) in first log differences. Column (3) includes state fixed effects. Robust standard errors clustered at the county level. *** p<0.01, ** p<0.05, * p<0.1

Table 2: Estimates of path dependence: main

	(1) coal capacity	(2) new coal capacity	(3) non-coal capacity	(4) new non-coal capacity	(5) rel. coal capacity	(6) new rel. coal capacity
Relative effect of log dist. to obsolete mine (β^o_τ)						
2 decades lead	0.20	-0.87	-0.54	0.13	-0.69	-0.30
	[0.35]	[1.16]	[0.40]	[0.67]	[0.71]	[0.94]
1 decade lead	-0.056	-0.68	-0.65**	-0.55	-0.19	-0.53
	[0.29]	[0.96]	[0.29]	[0.77]	[0.47]	[1.00]
	–	–	–	–	–	–
1 decade lag	0.17	-0.65	0.010	-0.72	-0.17	-2.11
	[0.44]	[0.95]	[0.25]	[0.51]	[0.82]	[1.75]
2 decades lag	-0.64***	-2.36*	0.25	-0.19	-3.11***	-2.56*
	[0.19]	[1.40]	[0.28]	[0.54]	[0.82]	[1.45]
3 decades lag	-0.83***	-3.38***	0.076	-0.96	-3.04***	-4.21***
	[0.21]	[1.29]	[0.35]	[0.68]	[0.59]	[1.40]
4 decades lag	-1.03***	-2.74*	0.24	-0.89	-2.78***	-2.69*
	[0.25]	[1.51]	[0.37]	[0.90]	[0.55]	[1.52]
5 decades lag	-1.08***	-0.090	0.81*	0.44	-3.58***	-1.35
	[0.37]	[1.29]	[0.42]	[0.87]	[0.74]	[1.27]
6 decades lag	-0.98***	-0.55	0.71*	-0.65	-2.68***	-1.16
	[0.37]	[1.29]	[0.41]	[0.82]	[0.55]	[1.33]
7 decades lag	-2.71***	-4.61***	0.55	-0.76	-4.96***	-6.74***
	[0.72]	[1.47]	[0.61]	[1.04]	[1.11]	[1.84]
8 decades lag	-2.77***	-3.59**	0.89	0.15	-6.05***	-3.86**
	[0.73]	[1.68]	[0.72]	[1.48]	[1.31]	[1.97]
9 decades lag	-2.80***	-2.52	0.53	-1.50	-6.03***	-2.75
	[0.78]	[2.04]	[0.82]	[0.93]	[1.29]	[2.31]
10 decades lag	-2.83***	-1.21	1.03	-0.70	-5.75***	-1.24
	[0.75]	[4.40]	[0.95]	[1.33]	[1.28]	[4.86]
Contemp effect of log dist. to nearest mine (β^c)	-1.08***	-1.59***	-0.32	-1.94**	-1.36***	-2.34***
	[0.37]	[0.55]	[0.53]	[0.90]	[0.49]	[0.65]
Observations	2,371	2,371	2,371	2,371	2,371	2,371
No. of counties	261	261	261	261	261	261

Notes: Estimates of β^o_τ and β^c from Equation 1. Each model includes event time dummies, contemporaneous distance to nearest mine, county fixed effects, and state-year fixed effects. Counties from benchmark sample (see Figure A.7). Outcome in Columns (1) and (2) is coal capacity and new coal capacity. Outcome in Columns (3) and (4) is non-coal capacity and new non-coal capacity. Outcome in Columns (5) and (6) is relative coal capacity and new relative coal capacity. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 3: Estimates of path dependence: robustness to specification

	(1)	(2)	(3)	(4)
	Outcome is relative coal capacity			
Relative effect of log dist. to obsolete mine (β^o_τ)				
2 decades lead	-0.23	-0.18	-0.69	0.11
	[0.51]	[0.53]	[0.71]	[0.84]
1 decade lead	0.24	0.26	-0.19	0.27
	[0.49]	[0.51]	[0.47]	[0.60]
	–	–	–	–
1 decade lag	0.17	0.017	-0.17	-0.14
	[0.30]	[0.29]	[0.82]	[0.70]
2 decades lag	-1.23***	-1.33***	-3.11***	-2.26***
	[0.48]	[0.48]	[0.82]	[0.62]
3 decades lag	-1.08**	-1.06*	-3.04***	-2.78***
	[0.51]	[0.56]	[0.59]	[0.71]
4 decades lag	-0.95	-0.97	-2.78***	-2.57***
	[0.63]	[0.72]	[0.55]	[0.74]
5 decades lag	-0.96**	-1.13***	-3.58***	-4.04***
	[0.45]	[0.43]	[0.74]	[1.00]
6 decades lag	0.70	0.32	-2.68***	-2.93***
	[0.44]	[0.37]	[0.55]	[0.72]
7 decades lag	-1.15	-1.91**	-4.96***	-4.92***
	[1.12]	[0.89]	[1.11]	[1.18]
8 decades lag	-1.41	-2.17***	-6.05***	-5.93***
	[1.07]	[0.83]	[1.31]	[1.28]
9 decades lag	-1.26	-2.00**	-6.03***	-5.97***
	[1.08]	[0.82]	[1.29]	[1.42]
10 decades lag	-1.20	-1.96**	-5.75***	-5.97***
	[1.08]	[0.84]	[1.28]	[1.45]
Observations	2,371	2,371	2,371	2,108
No. of counties	261	261	261	261
County fixed effect	Yes	Yes	Yes	Yes
Contemp. dist. to nearest mine	No	Yes	Yes	Yes
State-year fixed effects	No	No	Yes	Yes
County controls	No	No	No	Yes

Notes: Estimates of β^o_τ from Equation 1. Each model includes county fixed effects and event time dummies. Counties from benchmark sample (see Figure A.7). Additional controls are added cumulatively across columns. Column (2) adds contemporaneous distance to nearest mine. Column (3) adds state-year fixed effects and is the benchmark model from Column (5) of Table 2. Column (4) adds county total population, number of manufacturing establishments, and manufacturing employment. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Estimates of path dependence: robustness to sample restrictions

	(1)	(2)	(3)	(4)	(5)
	Outcome is relative coal capacity				
Relative effect of log dist. to obsolete mine (β^o_τ)					
2 decades lead	-0.69	-1.21	0.050	-0.78	
	[0.71]	[0.79]	[0.67]	[0.71]	
1 decade lead	-0.19	-0.57	0.065	-0.27	
	[0.47]	[0.58]	[0.45]	[0.48]	
	–	–	–	–	–
1 decade lag	-0.17	-0.82	0.0058	-0.17	-0.61*
	[0.82]	[0.87]	[0.78]	[0.69]	[0.35]
2 decades lag	-3.11***	-4.22***	-2.34***	-2.81***	0.29
	[0.82]	[0.93]	[0.70]	[0.79]	[0.50]
3 decades lag	-3.04***	-3.51***	-2.11***	-2.80***	-1.68***
	[0.59]	[0.65]	[0.57]	[0.57]	[0.62]
4 decades lag	-2.78***	-3.45***	-1.83***	-2.56***	-2.50***
	[0.55]	[0.68]	[0.56]	[0.54]	[0.87]
5 decades lag	-3.58***	-4.98***	-1.57**	-3.03***	-2.69***
	[0.74]	[1.10]	[0.62]	[0.63]	[1.01]
6 decades lag	-2.68***	-3.02***	-1.40**	-2.63***	-1.85***
	[0.55]	[0.77]	[0.63]	[0.51]	[0.72]
7 decades lag	-4.96***	-5.22***	-3.36***	-3.32***	-4.24***
	[1.11]	[1.12]	[1.03]	[0.70]	[1.22]
8 decades lag	-6.05***	-6.38***	-4.23***	-3.92***	-5.39***
	[1.31]	[1.26]	[1.21]	[0.95]	[1.38]
9 decades lag	-6.03***	-6.37***	-4.09***	-3.87***	-5.40***
	[1.29]	[1.20]	[1.19]	[0.96]	[1.34]
10 decades lag	-5.75***	-6.05***	-3.83***	-3.81***	-5.09***
	[1.28]	[1.19]	[1.19]	[0.95]	[1.34]
Observations	2,371	1,938	2,881	3,218	1,452
No. of counties	261	208	320	338	132
Sample	Benchmark	<200 miles from Ill. Basin	<300 miles from Ill. Basin	Include closer to App. Basin	Treated < 1960s

Notes: Estimates of β^o_τ from Equation 1. Each model includes event time dummies, contemporaneous distance to nearest mine, county fixed effects, and state-year fixed effects. Column (1) replicates benchmark model from Column (5) of Table 2. Columns (2) and (3) restricts sample to counties within 200 and 300 miles from the nearest Illinois Basin coal resource and are closer to Illinois Basin coal than to Appalachian Basin coal. Column (4) restricts sample to counties within 250 from the nearest Illinois Basin coal resource but include counties that are closer to Appalachian Basin coal than to Illinois Basin coal. Column (5) restricts benchmark sample to include only counties experiencing event prior to the 1960s. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: Estimates of path dependence: robustness to data construction

	(1)	(2)	(3)	(4)	(5)
	Outcome is relative coal capacity				
Relative effect of log dist. to obsolete mine (β^o_τ)					
2 decades lead	-0.69	-1.07	-0.46	-0.82	0.28
	[0.71]	[0.84]	[0.70]	[0.70]	[0.90]
1 decade lead	-0.19	0.019	0.70	-0.24	0.22
	[0.47]	[0.61]	[0.79]	[0.46]	[0.57]
	–	–	–	–	–
1 decade lag	-0.17	-0.054	-0.22	-0.22	0.35
	[0.82]	[0.91]	[1.00]	[0.78]	[0.99]
2 decades lag	-3.11***	-3.23***	-3.32***	-3.03***	-1.87
	[0.82]	[0.87]	[0.92]	[0.76]	[1.16]
3 decades lag	-3.04***	-3.01***	-3.22***	-3.40***	-1.72*
	[0.59]	[0.61]	[0.67]	[0.66]	[1.00]
4 decades lag	-2.78***	-2.86***	-3.02***	-3.28***	-1.75
	[0.55]	[0.56]	[0.62]	[0.67]	[1.19]
5 decades lag	-3.58***	-3.68***	-3.92***	-4.03***	-2.86**
	[0.74]	[0.75]	[0.79]	[0.85]	[1.22]
6 decades lag	-2.68***	-2.69***	-2.85***	-3.18***	-1.75
	[0.55]	[0.60]	[0.68]	[0.68]	[1.21]
7 decades lag	-4.96***	-4.94***	-5.60***	-5.41***	-3.37**
	[1.11]	[1.05]	[1.23]	[1.24]	[1.50]
8 decades lag	-6.05***	-6.05***	-6.58***	-6.34***	-4.14**
	[1.31]	[1.27]	[1.43]	[1.35]	[1.69]
9 decades lag	-6.03***	-6.06***	-6.62***	-6.38***	-4.63***
	[1.29]	[1.21]	[1.36]	[1.32]	[1.62]
10 decades lag	-5.75***	-5.78***	-6.38***	-5.59***	-4.49***
	[1.28]	[1.19]	[1.36]	[1.38]	[1.60]
Observations	2,371	2,325	2,264	2,034	2,080
No. of counties	261	261	261	227	228
Num of nearest mines	1	2	3	1	1
EIA data year	1990-2012	1990-2012	1990-2012	2000	2005

Notes: Estimates of β^o_τ from Equation 1. Each model includes event time dummies, contemporaneous distance to nearest mine, county fixed effects, and state-year fixed effects. Counties from benchmark sample (see Figure A.7). Column (1) shows benchmark model. Column (2) uses average distance to 2 nearest mines. Column (3) uses average distance to 3 nearest mines. Outcome in Column (4) built from only the 2000 EIA 860 form. Outcome in Column (5) built from only the 2005 EIA 860 form. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: Estimates of path dependence: robustness to other modeling choices

	(1)	(2)	(3)	(4)	(5)
	Outcome is relative coal capacity				
Relative effect of log dist. to obsolete mine (β^o_τ)					
2 decades lead	-0.69	-0.69	-2.66	-0.15	0.36
	[0.71]	[0.78]	[2.22]	[0.35]	[1.94]
1 decade lead	-0.19	-0.19	-1.53	0.26	0.87
	[0.47]	[0.44]	[0.94]	[0.29]	[1.11]
	–	–	–	–	–
1 decade lag	-0.17	-0.17	-0.84	-0.31	-0.66**
	[0.82]	[0.90]	[0.58]	[0.24]	[0.27]
2 decades lag	-3.11***	-3.11***	-2.49	-0.92***	-0.88**
	[0.82]	[0.74]	[2.02]	[0.25]	[0.42]
3 decades lag	-3.04***	-3.04***	-4.18**	-1.35***	-0.94
	[0.59]	[0.64]	[2.03]	[0.41]	[0.57]
4 decades lag	-2.78***	-2.78***	-4.97**	-1.39***	-0.91
	[0.55]	[0.64]	[1.94]	[0.44]	[0.67]
5 decades lag	-3.58***	-3.58***	-3.91*	-1.74***	-1.61*
	[0.74]	[0.94]	[2.08]	[0.54]	[0.82]
6 decades lag	-2.68***	-2.68***	-2.82	-1.43**	-1.69*
	[0.55]	[0.73]	[2.20]	[0.69]	[0.87]
7 decades lag	-4.96***	-4.96***	0.71	-2.33**	-0.077
	[1.11]	[1.17]	[2.96]	[1.08]	[1.21]
8 decades lag	-6.05***	-6.05***	-3.33	-3.30***	0.026
	[1.31]	[1.22]	[2.64]	[1.13]	[1.18]
9 decades lag	-6.03***	-6.03***	-2.63	-3.09**	0.48
	[1.29]	[1.19]	[2.45]	[1.26]	[1.41]
10 decades lag	-5.75***	-5.75***	-2.2	-3.20**	1.56
	[1.28]	[1.23]	[2.83]	[1.40]	[2.14]
Observations	2,371	2,369	382	2,371	298
No. of counties	261	261	66	261	71
Model	Poisson	Poisson	Poisson	Linear	Linear
Add 1?	Yes	Yes	No	Yes	No
Standard error clustering	County	State-year	County	County	County

Notes: Estimates of β^o_τ from Equation 1. Each model includes event time dummies, contemporaneous distance to nearest mine, county fixed effects, and state-year fixed effects. Counties from benchmark sample (see Figure A.7). Robust standard errors clustered at the county-level for all models except Column (2). Column (1) replicates benchmark model from Column (5) of Table 2. Column (2) uses state-year clustered standard errors. Column (3) uses unadjusted relative coal capacity as outcome. Column (4) estimates a linear model on adjusted relative coal capacity. Column (5) estimates a linear model on unadjusted relative coal capacity. *** p<0.01, ** p<0.05, * p<0.1.

Table 7: Returns to scale regressions at the power plant level

	(1)	(2)	(3)	(4)	(5)	(6)
	Dep. var. is log non-fuel cost					
log elec output (mwh)	0.56***	0.48***	0.60***	0.52***	0.57***	0.51***
	[0.028]	[0.035]	[0.087]	[0.072]	[0.052]	[0.054]
CLR P-value			0.0081	0.0075		
CLR confidence int (90%)			[.54,.89]	[.44, .76]		
implied scale parameter, ψ	1.78***	2.09***	1.66***	1.93***	1.75***	1.96***
	[0.090]	[0.15]	[0.24]	[0.27]	[0.16]	[0.21]
Model	OLS	OLS	IV	IV	OLS	OLS
County sample	Benchmark	Benchmark	Benchmark	Benchmark	U.S.	U.S.
Fuel input	Only coal	Primarily coal	Only coal	Primarily coal	Primarily gas	Primarily oil
No. of power plants	103	147	96	139	32	73

Notes: Estimates from Eq. 6 and 7. Each model includes average (1981-1999) observed power plant level fuel price, state and NERC region fixed effects, and county centroid longitude and latitude. Columns (1)-(4) includes power plants in benchmark sample (see Figure A.7). OLS regressions for power plants burning only coal and primarily coal shown in Columns (1) and (2) respectively. IV regressions for power plants burning only coal and primarily coal shown in Columns (3) and (4) respectively. OLS regressions in Columns (5) and (6) include all U.S. power plants burning primarily natural gas and oil. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Productivity regressions at the generator level

	(1)	(2)	(3)	(4)
	Dep. var. is log generator thermal efficiency			
$\ln dist_i^o X sinceEvent_i$	-0.012	-0.0083	-0.0027	-0.0074
	[0.014]	[0.010]	[0.0100]	[0.0070]
$\ln dist_i^o$	0.06	0.038	0.012	0.015
	[0.054]	[0.045]	[0.054]	[0.039]
$sinceEvent_i$	0.022	0.021	-0.008	0.027
	[0.058]	[0.039]	[0.039]	[0.027]
age		-0.0057**		-0.0062***
		[0.0022]		[0.0018]
capacity		0.000062		0.000082
		[0.000084]		[0.000068]
Fuel input	Only coal	Only coal	Primarily coal	Primarily coal
No. of generators	160	160	266	266
No. of power plants	63	63	96	96

Notes: Estimates from Eq. 8. All models includes county distance to nearest coal mine, state and NERC region fixed effects, and county centroid longitude and latitude. Includes power plants from benchmark county sample (see Figure A.7). Columns (1) and (2) include coal-fired generators in plants that only burn coal. Columns (3) and (4) include coal-fired generators in plants that primarily burn coal. Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Table 9: Examining electricity demand-side mechanisms

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Decades since obsolescence							
Outcome	1	2	3	4	5	6	7	8
Log population	-0.047***	-0.063**	-0.052**	-0.03	-0.00093	0.028	-0.019	-0.0012
	[0.016]	[0.025]	[0.022]	[0.036]	[0.041]	[0.038]	[0.031]	[0.030]
No. of counties	261	231	229	132	119	107	107	107
Log urban population	-0.025	-0.00012	-0.022	-0.12	0.15***	0.16***	-0.04	-0.094
	[0.050]	[0.054]	[0.046]	[0.11]	[0.056]	[0.055]	[0.096]	[0.061]
No. of counties	153	164	92	89	84	88	93	97
log mgf establishments	0.0019	-0.019	-0.051	-0.17**	0.0085	0.078	0.0099	0.11*
	[0.032]	[0.048]	[0.056]	[0.066]	[0.055]	[0.097]	[0.066]	[0.066]
No. of counties	154	125	221	123	117	107	107	107
log mgf employment	-0.018	0.037	0.08	0.093	0.031	-0.19	0.11	-0.092
	[0.070]	[0.063]	[0.089]	[0.10]	[0.16]	[0.14]	[0.12]	[0.14]
No. of counties	139	103	193	114	107	104	95	87

Notes: Each coefficient from a separate regression showing effect of distance to obsolete coal mine on county characteristics since obsolescence. Outcomes in log first differences. All models include state fixed effects. Counties from benchmark sample (see Figure A.7). Robust standard errors clustered at the county level. *** p<0.01, ** p<0.05, * p<0.1

Table 10: Effect of distance to obsolete coal mine on modern transport density

	(1)	(2)
	Rail density (2010)	Highway density (2010)
$\ln dist_i^o X sinceEvent_i$	-0.031	-0.027
	[0.031]	[0.018]
$\ln dist_i^o$	0.039	0.0021
	[0.17]	[0.11]
$sinceEvent_i$	0.13	0.075
	[0.13]	[0.073]
No. of counties	458	458

Notes: Estimates from cross-sectional specification as in Eq. 8. Outcomes in logs. Each model includes county-level distance to nearest coal mine, state and NERC region fixed effects, and county centroid longitude and latitude. Counties from benchmark sample (see Figure A.7). Robust standard errors clustered at the county-level. *** p<0.01, ** p<0.05, * p<0.1.

Appendix to

Estimating Path Dependence

in Energy Transitions

FOR ONLINE PUBLICATION

A Data Sources

This section details the data used in the paper.

A.1 Coal resources, mining, and modern delivered coal prices

U.S.G.S. National Coal Resource Assessment (N.C.R.A.) This paper uses two sets of spatial data from the N.C.R.A. (East, 2012). The first dataset contains vector-based shape files of Illinois and Appalachian Basin coal resources that are situated less than and above 200 feet from the surface. These shape files are used to generate the map in Figure A.7 showing each depth layer of coal resources for the two basins. The second dataset contains characteristics of all coal mines over the Illinois Basin that has operated since 1890. Variables include geo-coded location, opening year, closing year, and spatial area of the mine. Because the coal mines database does not include the depth of the coal resource being mined, this dataset is spatially overlaid onto the depth-specific coal resource shape files, as shown Figure 4, and used to generate a county-decade panel of delivered coal prices displayed in Figure A.8. The timing of treatment for each sample county is shown in Figure A.9

EIA 423 forms (1972-1999) Verification of my constructed county delivered coal price panel dataset is performed against coal procurement data available from 1972-1999 EIA 423 forms. The raw data is at the destination power plant by county of coal origin level. It contains data on the quantity and price of the purchased coal as well as its heat, sulfur, and ash content. EIA notes that the county of coal origin may be unreliable.[44] A spot check of the data shows this to be particularly the case for EIA 423 forms before 1990 as the county identifier recorded does not match standard county identifiers.

This data is used to serve three objectives. First, using the 1990-1999 period where coal county of origin is more reliable, EIA 423 data on coal heat, sulfur, and heat content is aggregated to the county of origin and averaged across the years to produce the maps in Figure A.6 documenting relative homogeneity in coal quality in the Illinois Basin. Second, data from 1990-1999 EIA 423 forms are used to test underlying assumptions in the construction of delivered coal prices via the Hotelling local model as discussed in Section 4.1 and shown in Figure A.11. Finally, because the county of the destination power plant is recorded relatively accurate, I use the entire set of 1972-1999 EIA 423 forms to construct decade-by-county level delivered coal prices to verify against my constructed delivered coal prices. This is used in the regression results in Table A.1.

A.2 Electricity production data

Generator-level data: EIA 860 forms (1980, 1990-2012) This paper uses EIA-860 data in three days. First, fuel-specific capacity, county of location, opening year, and closing year is used to construct the county-by-decade-by-fuel capacity panel dataset from 1890-1990. The assumptions needed for this data construction, discussed in Section 4.2 are tested in Figure A.14 and Tables A.2 and A.3 using both 1990-2012 EIA 860 forms as well as the 1980 EIA 860 form, which was digitized for this paper. This data was also used when comparing with historical national and state-level data on generating capacity as shown in Figures A.12 and A.13. Second, this constructed panel of historical county-by-decade-by-fuel capacity is used to

[44] In the EIA 423 readme file, it is noted that "The instructions for the FERC Form 423 require the respondent to report the county in which the coal was mined. However, this data is not always known or reported correctly. Sometimes a utility will report the location of a tipple or a prep plant as the source of the coal. In these cases, the coal usually comes from the surrounding counties. In some instances, a utility may report the county where MOST of the coal came from. It is very difficult to verify county level data. Users of the data should be aware of this and use the data accordingly."

generate the prima-facie evidence in Figures 1 and 2. Third, the 1990-1995 EIA 860 forms include data on generator heat rate, which is the inverse of thermal efficiency.[45] I use log average thermal efficiency during 1990-1995 as the outcome variable in Table 8.

Power plant-level data: Historical FERC map and directories (1941, 1945, 1951, 1954, 1963) To verify my county-decade-fuel capacity dataset constructed from modern EIA 860 forms, I construct a state-decade-prime mover panel using the historical power plant dataset constructed by Lewis (2014). This dataset is constructed by first digitizing a map of power plants in 1963 (Federal Power Commission, 1963) which is then merged with power plant the timing of each power plant opening from directories on electric generating plants (Federal Power Commission, 1941, 1945, 1951, 1954). See Lewis (2014) for more details. This data is used in Figure A.13.

Power plant-level data: PLATTS/UDI O&M cost dataset (1981-1999) Power plant-level data on total and fuel costs for 1981-1999 was purchased from PLATTS/UDI. It is used to construct log non-fuel cost, the outcome variable in Table 7.

U.S. time series data (1890-2012) This paper uses two sources of aggregate U.S. time series data. U.S. Census Bureau (1975) provides total and mechanically produced U.S. bituminous coal production from 1890-1950 (see Figure A.4) and total electricity capacity by steam and hydropower from 1920-1970 (see Figure A.12). For the share of national delivered coal prices due to transport costs from 1902-2007), I use data from McNerney, Farmer and Trancik (2011). National coal and natural gas sales price from Energy Information Administration (2012) (see Figure A.16).

Cross country data Country-level per capita CO_2 emissions and GDP per capita in 1960 and 2000 come from Boden and Andres (2013) and World Bank (2014), respectively. Shown in Figure A.1.

A.3 Other

County-level U.S. census data (1890-1990) County-level historical variables from 1890-2000 are obtained from historical U.S. censuses as collected by Haines and Inter-university Consortium for Political and Social Research (2010).[46] These variables include total population, rural population as well as manufacturing variables such as number of establishments, employment, capital value, and output. Because U.S. county boundaries were evolving until the 1930s, I redraw county-level data from 1890 to 1930 onto 1930 county boundaries using historical GIS county shape files from the U.S. National Historical Geographical Information System (N.H.G.I.S.)[47] using a method which modifies the procedure developed by Hornbeck (2012). This data is used to test for pre-trends in Table 1, as controls in my path dependence specification in Table 3, and in tests of alternative mechanisms in Table 9.

County-level transportation data (2010) Modern county-level data on highways and railroad networks was obtained from the U.S. Department of Transportation's National Transportation Atlas Database.[48]

[45]Specifically, efficiency is 3412 divided by the heat rate where 3412 is the equivalent Btu content of a kWh of electricity. See `https://www.eia.gov/tools/faqs/faq.cfm?id=107&t=3` for more details.

[46]Available: `http://doi.org/10.3886/ICPSR02896.v3`

[47]Available: `https://www.nhgis.org/documentation/gis-data`

[48]Available: `http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2012/index.html`

These variables are used as outcomes in the regressions shown in Table 10.

B Historical data on coal prices and electricity capacity and their current availability

Sections 4.1 and 4.2 mentioned that actual historical data on delivered coal prices and fuel-specific electricity capacity at or below the county-level from 1890-1990 were either never collected or if collected may no longer exist. This section summarizes the data that was historically collected, whether it is relevant for this study, and whether it is available today.

B.1 Coal prices

1882-1970

County coal producer prices are recorded in the U.S. Geological Survey's Bureau of Mine reports "Mineral resources of the United States 1882-1931" and "Minerals yearbook 1932-1970". However, the mapping from county-level producer to delivered coal price is not straightforward as a mine in one county can supply multiple counties.

Availability Available online.[49]

B.2 Electricity capacity

1902-1917:

The U.S. federal government first collected power plant-level data in 1902 in the inaugural Central Electric Light and Power Station Census administered by the then Department of Commerce and Labor. This was repeated every five years in 1907, 1912, and 1917. Importantly, power plants are classified by prime-mover (i.e. steam, hydro, internal combustion) and not fuel input needed for this paper. Input costs are provided for each fuel. However, in order to recover an input-specific capacity, fuel price and a generator capacity factor would be needed, neither of which were recorded

Availability Summaries of these censuses with aggregate statistics are available digitally.[50] However, the original power plant-level data could not be located following private conversations with archivists at the National Archives and Records Administration (NARA).[51]

1920-1970:

Soon after the creation of the Federal Power Commission (FPC) in 1920, several forms were administered annually to document electricity production and fuel consumption. The most important of these forms were the Annual Financial and Statistical Reports (Form 1) and the Power System Statements (Form 12). Form 1 shows generator capacity and built year but not type of input fuel. Form 1 shows primary fuel usage by

[49] Available: `hathitrust.org`

[50] Available: `http://hdl.handle.net/2027/mdp.39015028113663`

[51] Once a document has exceeded its agency retention period (which pertains to all pre-1970 documents), only 3% of documents are deemed permanently valuable and retained in the public collection of at NARA.

fuel type at the power-plant level. This data has two limitations. First, in order to recover plant-by-input fuel capacity, one needs plant-by-input fuel capacity factors, which is not available. Second, power plant coverage is incomplete. In 1948, Forms 1 and 12 included only 219 power plants covering 67% of total U.S. steam capacity.

Availability State-level aggregate statistics for electricity capacity by prime-mover and energy input consumption by fuel is available in the annual "Production of Energy and Capacity of Plants and Fuel Consumption of Electric Power Plants" and compiled in the "Electric Power Statistics, 1920-1940".[52] The "Steam-Electric Plant Construction Cost and Annual Production Expenses" from 1948 to 1974 has plant-level values from Form 1 and Form 12.[53] This data was also printed on a 1963 map called "Principal electric power facilities in the United States" and digitized by Lewis (2014) and used in the data validation exercises in Section 4.2. Private conversations with NARA archivists concluded that NARA may no longer hold the original Form 4 and Form 12 documents.

1977-1990:

The FPC was replaced by the Federal Energy Regulatory Commission (FERC) under the U.S. Department of Energy in 1977. FERC began publishing the "Inventory of Power Plants in the United States" that year combining generator data from the Monthly Power Plant Report (Form 4), Annual Power System Statement (Form 12), and the Supplemental Power Statement (Form 12E-2). This report includes generator capacity, type of energy input, and built year. Unfortunately, retired units documented in the "Inventory of Power Plants in the United States" only included units that were retired during record year and not previously retired generators.

Availability The "Inventory of Power Plants in the United States" for 1980 is available online[54] and was digitized for my data validation exercises discussed in Section 4.2. Reports from other years are kept as microfiche in many research libraries.

1985 - :

The Annual Electric Generator Report, Form EIA-860, was originally implemented in 1985. It replaced the previous Form 4, Form 12 and 12E, Form 67, and Form 411. According to the U.S. Energy Information Agency which administers the form:

> "The Form EIA-860 is a mandatory annual census of all existing and planned electric generating facilities in the United States with a total generator nameplate capacity of 1 or more megawatts. The survey is used to collect data on existing power plants and 10 year plans for constructing new plants, as well as generator additions, modifications, and retirements in existing plants. Data on the survey are collected at the individual generator level."

Availability Available online.[55]

[52] Available: `http://hdl.handle.net/2027/mdp.39015023906806`
[53] Available: `http://catalog.hathitrust.org/Record/000904499`
[54] Available: `http://hdl.handle.net/2027/umn.31951d02987924n`
[55] Available: `http://www.eia.gov/electricity/data/eia860/`

C Electricity producer problem

C.1 Profit maximization

Inputs for older vintaged generators are fixed. In period t, the power plant chooses inputs only for current-vintage generators. The optimization problem is:

$$\max_{X_{ct},X_{nt},E_{ct},E_{nt}} \left(Y_{ct}^{(\epsilon-1)/\epsilon}+Y_{nt}^{(\epsilon-1)/\epsilon}\right)^{\epsilon/(\epsilon-1)} - z_{ct}(E_{ct}+E_{ct-1}) - z_{nt}(E_{nt}+E_{nt-1}) - r_t(X_{ct}+X_{nt})$$

where z_{jt} is the primary energy input price and r_t is the rental rate of capital. Inserting the intermediate good production function from Eq. 4 and observing that efficient input implies $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ and $A_{Xjt-1}\delta X_{jt-1} = A_{Ejt-1}E_{jt-1}$ for each vintage- and fuel-specific generator, the problem can be written entirely in terms of capital:

$$\begin{aligned}\max_{X_{ct},X_{nt}} &\left((A_{Xct}X_{ct}A_{Xct-1}\delta X_{ct-1})^{\alpha(\epsilon-1)/\epsilon}+(A_{Xnt}X_{nt}A_{Xnt-1}\delta X_{nt-1})^{\alpha(\epsilon-1)/\epsilon}\right)^{\epsilon/(\epsilon-1)}\\ &-z_{ct}(\frac{A_{Xct}}{A_{Ect}}X_{ct}+\frac{A_{Xct-1}}{A_{Ect-1}}X_{ct-1})-z_{nt}(\frac{A_{Xnt}}{A_{Ent}}X_{nt}+\frac{A_{Xnt-1}}{A_{Ent-1}}X_{nt-1})-r_t(X_{ct}+X_{nt})\end{aligned} \tag{A.1}$$

Suppose input-specific capital productivities are the same across intermediate goods during period $t-1$, $A_{Xct-1} = A_{Xnt-1}$ and $A_{Ect-1} = A_{Ent-1}$. Taking the ratio of the two first order conditions for Eq. A.1 and rewriting in terms of relative current-vintage coal capacity, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$:

$$\widetilde{X}_t = \widetilde{w}_t^{\frac{\epsilon}{\varphi-1}}\,\widetilde{X}_{t-1}^{\frac{\alpha(1-\epsilon)}{\varphi-1}}\,\widetilde{A}_{Xt}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \tag{A.2}$$

where $w_{jt} = \frac{A_{Xjt}}{A_{Ejt}}z_{jt} + r_t$ is the productivity weighted input price index and $\varphi = (1-\alpha)(1-\epsilon) < 0$. Eq. A.2 is Eq. 5 in the main text.

C.2 Cost minimization problem to recover scale parameter

Consider a power plant m containing only coal-fired generators such that one can drop index j and only needs to consider the intermediate good production function Eq. 4. Applying fixed input proportions $E_{mt} = \frac{A_{Xmt}}{A_{Emt}}X_{mt}$ and $E_{mt-1} = \frac{A_{Xmt-1}}{A_{Emt-1}}X_{mt-1}$, the constrained cost minimization problem can be written in terms of fuel inputs:

$$\begin{aligned}C(z_{mt},r_{mt},Y_{mt}) = \min_{E_{mt}} \quad & z_{mt}\,(E_{mt}+E_{mt-1})+r_t(\frac{A_{Emt}}{A_{Xmt}\delta}E_{mt})\\ s.t. \quad & Y_{mt} = (A_{Emt}E_{mt}A_{Emt-1}E_{mt-1})^{\alpha}\end{aligned}$$

Rewriting the production function as $E_{mt-1} = Y_{mt}^{1/\alpha}(A_{Emt}A_{Emt-1}E_{mt})^{-1}$, one obtains the following equivalent unconstrained minimization problem:

$$\min_{E_{mt}} \quad (z_{mt}+\frac{A_{Emt}}{A_{Xmt}\delta}r_t)E_{mt}+z_{mt}Y_{mt}^{1/\alpha}(A_{Emt}A_{Emt-1}E_{mt})^{-1} \tag{A.3}$$

Taking the first order condition of Eq. A.3 yields a conditional demand function:

$$E^*_{mt} = (Y_{mt})^{1/2\alpha}(\frac{z_{mt}}{z_{mt}+\frac{A_{Emt}}{A_{Xmt}\delta}})^{1/2}(A_{Emt}A_{Emt-1})^{-1/2} \tag{A.4}$$

Inserting Eq. A.4 into non-fuel cost at the cost-minimizing input level, $\text{non\_fuel\_cost}_{mt} = C(z_{mt}, r_{mt}, Y_{mt}) - z_{mt}(E^*_{mt} + E_{mt-1}) = r_t(\frac{A_{Emt}}{A_{Xmt}\delta} E^*_{mt})$, and applying a natural logs yield:

$$\ln(\text{non\_fuel\_cost}_{mt}) = \frac{1}{\psi}\ln Y_{mt} + \frac{1}{2}\ln(\frac{z_{mt}}{z_{mt} + \frac{A_{Emt}}{A_{Xmt}\delta}}) + \ln r_t + \ln(A^{1/2}_{Emt} A^{-1/2}_{Emt-1} A^{-1}_{Xmt} \delta^{-1}) \tag{A.5}$$

where $\psi = 2\alpha$. Eq. A.5 is the structural analog to the OLS specification in Eq. 6 from the main text.

D Simulating future emissions

This section details the procedure for simulating future CO_2 emissions following a relative coal price shock, as shown in Figure 7.

D.1 Parameters

- Estimated returns to scale parameter, mean $\hat{\psi} = 1.66$, standard error $\hat{\sigma}_\psi = 0.24$.
- Estimated new relative capacity effect (one generation), mean $\widehat{\beta^o} = -2.84$, standard error $\hat{\sigma}_{\beta^o} = 1.17$.
- Baseline relative coal prices set to 2000's value, $\widetilde{w}_t = 0.4$.
- Relative coal price shock based on 98% increase in relative coal prices during 2009-2010 (see Figure A.16).
- Relative productiviy set to 2000's value, $\widetilde{A}_{Xt} = 0.7$.
- Capital depreciation rate set to 2000's value, $\delta = 0.06$.
- Carbon content of coal: $C_c = 4931.3$ lb CO_2/short ton coal.[56]
- Carbon content of natural gas: $C_n = 119.9$ lb CO_2/ thousand cubic feet.[57]

D.2 Historical emissions

After aggregating coal and natural gas capacity to the national level, emissions is calculated by:

$$M_t = E_{ct}C_c + E_{nt}C_n$$

where E_{ct} and E_{nt} is total coal (in short tons) and natural gas (in thousand cubic feet) consumed respectively by the U.S. electricity sector in year t. Data from Energy Information Administration (2012).

D.3 Simulating future emissions

Set relative coal price, $\widetilde{w}_t$, for each decade $t = 2010...2150$. Under baseline case, $\widetilde{w}_t = 0.4\ \forall t$. Under shock experiment with duration $d = 10, 20, 30$ and shock multiplier $M = 1, 1.5, 3$, $\widetilde{w}_t = 0.4 + 0.98M * \mathbf{1}(t = 2010 + d)$.

Conduct the following Monte Carlo procedure with $b = 1...250$ draws:

[56] Available here: `https://www.eia.gov/environment/emissions/co2_vol_mass.cfm`
[57] Available here: `https://www.eia.gov/environment/emissions/co2_vol_mass.cfm`

- Draw $\psi(b) \sim N(\hat{\psi}, \hat{\sigma_\psi})$ and $\beta^o(b) \sim N(\widehat{\beta^o}, \hat{\sigma}_{\beta^o})$. Define $\alpha(b) = \psi(b)/2$.
- Obtain $\epsilon(b)$ by taking the positive root from:

  $$\epsilon(b)^2[2(1-\beta^o(b))\alpha(b)-(1-\beta^o(b))+\beta^o(b)\alpha(b)^2]+\epsilon(b)[-2(1-\beta^o(b))\alpha(b)-2\beta^o(b)\alpha(b)^2]+\beta^o(b)\alpha(b)^2 = 0$$

  Define $\varphi(b) = (1-\alpha(b))(1-\epsilon(b))$
- For each decade $t = 2010...2150$, implement:

  1. Apply Eq. 9 to obtain new relative coal capacity:

  $$\widetilde{\Delta X}(b)_t = exp\left(\sum_{s=0}^{\infty}\left(\frac{\epsilon(b)}{(\psi(b)-1)}\left[\frac{\alpha(b)(1-\epsilon(b))}{(\varphi(b)-1)}\right]^s \ln \widetilde{w}_{t-s} + \left[\frac{\alpha(b)(1-\epsilon(b))}{(\varphi(b)-1)}\right]^{s+1} \ln \widetilde{A}_{t-s}\right)\right)$$

  2. Obtain new coal capacity, holding total capacity constant, less depreciation, from previous period:

  $$\Delta X(b)_{ct} = \left(\frac{1}{\frac{1}{\Delta\widetilde{X}(b)_t}+1}\right)(X(b)_{ct-1} + X(b)_{nt-1})\,\delta$$

  3. Obtain new natural gas capacity, holding total capacity constant from previous period:

  $$\Delta X(b)_{nt} = \left(1 - \frac{1}{\frac{1}{\Delta\widetilde{X}(b)_t}+1}\right)(X(b)_{ct-1} + X(b)_{nt-1})\,\delta$$

  4. Obtain cumulative coal capacity:

  $$X(b)_{ct} = X(b)_{ct-1}(1-\delta) + \Delta X(b)_{ct}$$

  5. Obtain cumulative natural gas capacity:

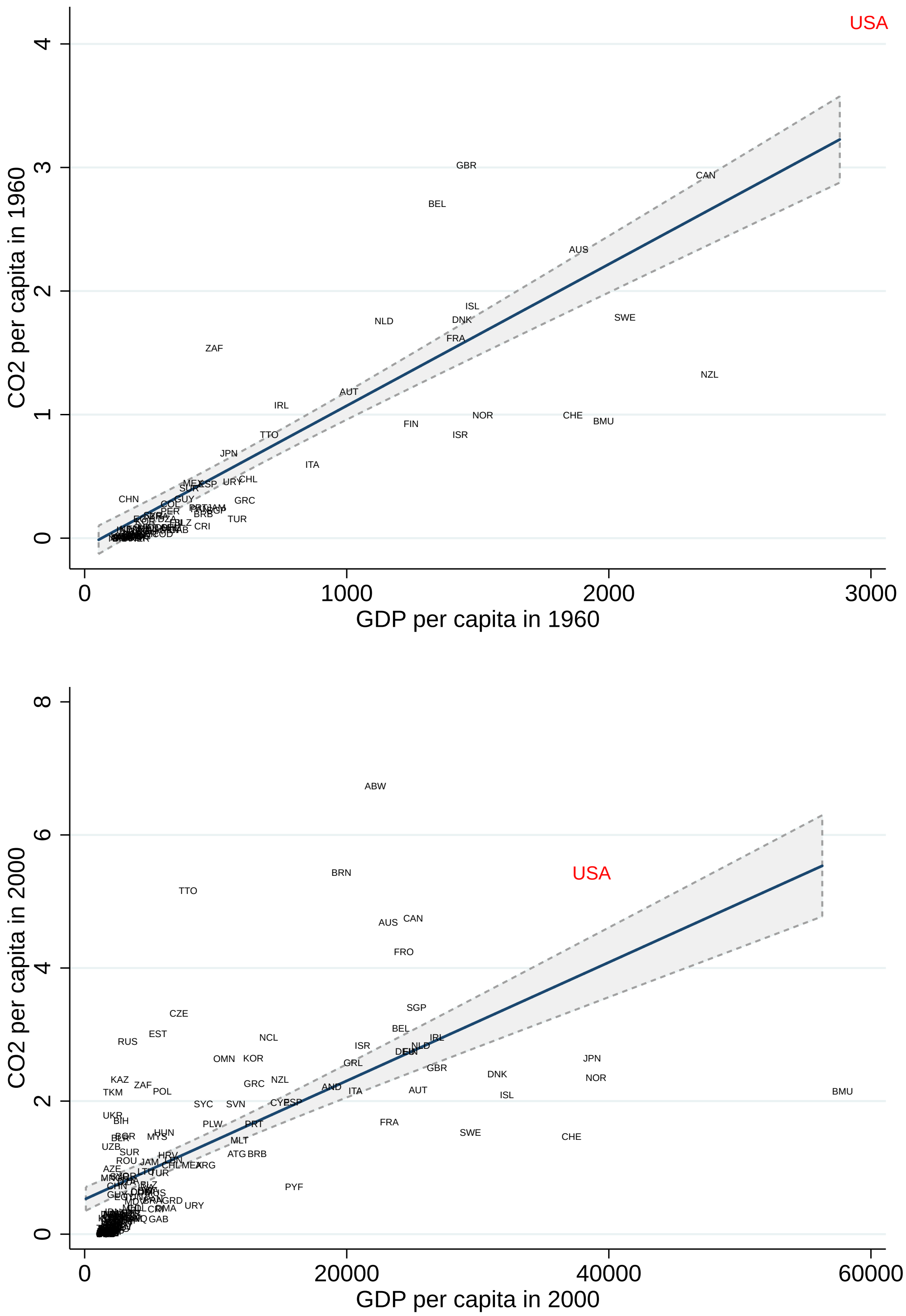
  $$X(b)_{nt} = X(b)_{nt-1}(1-\delta) + \Delta X(b)_{nt}$$

  6. Obtain total CO_2 emissions using 2010 emissions intensity:

  $$M(b)_t = X(b)_{ct}\frac{E_{c2010}}{X_{c2000}}C_c + X(b)_{nt}\frac{E_{n2010}}{X_{n2000}}C_n$$

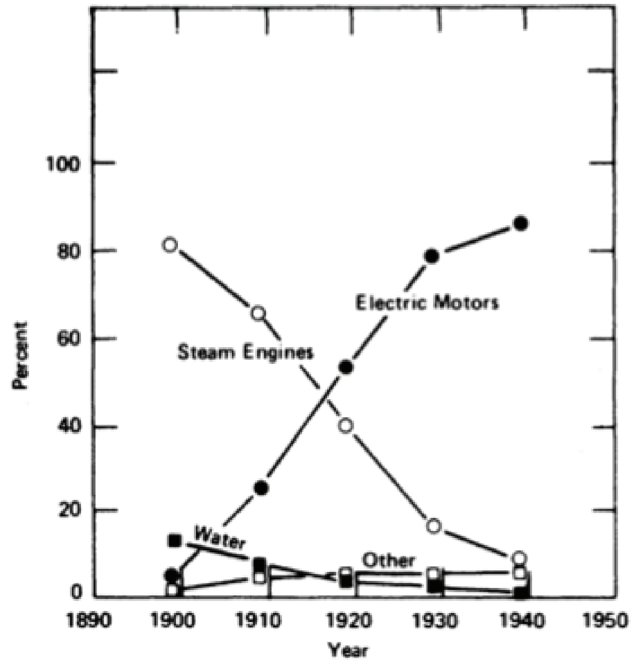
Appendix Figures

Figure A.1: U.S. CO_2 emissions intensity in 1960 and 2000



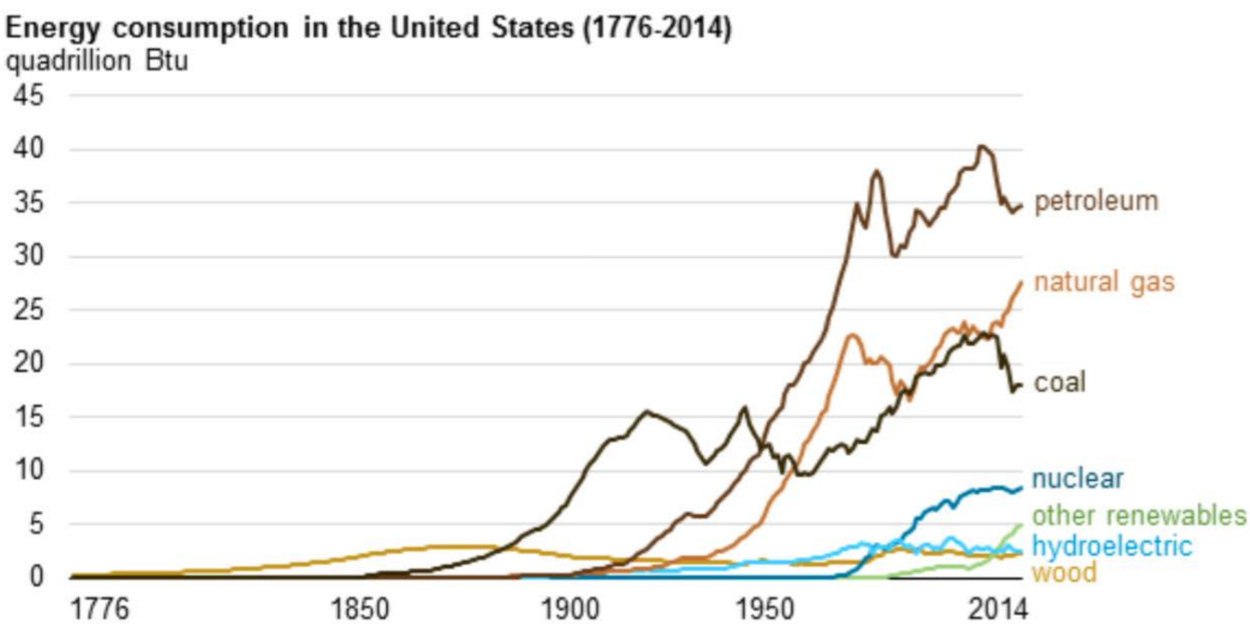
Notes: Scatter shows carbon dioxide emissions per capita against income per capita in 1960 (top panel) and 2000 (bottom panel). Linear regression fit shown with 90% confidence interval. OPEC countries excluded. Source: Boden and Andres (2013) and World Bank (2014).

Figure A.2: Historical U.S. intermediate energy shares



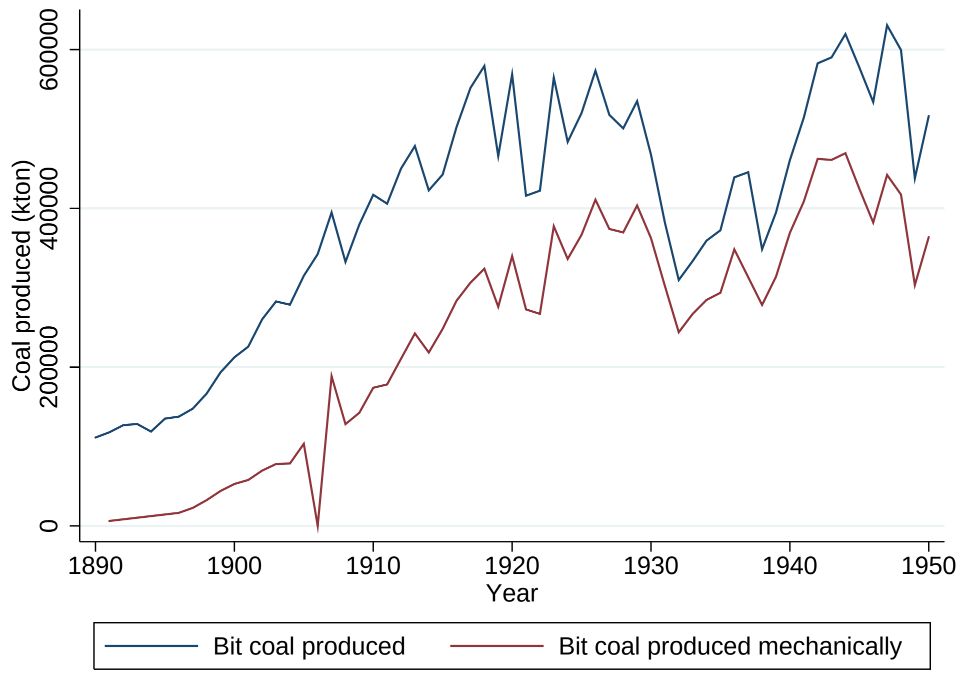
Notes: Reproduced from Devine (1983).

Figure A.3: U.S. energy conusmption by primary input



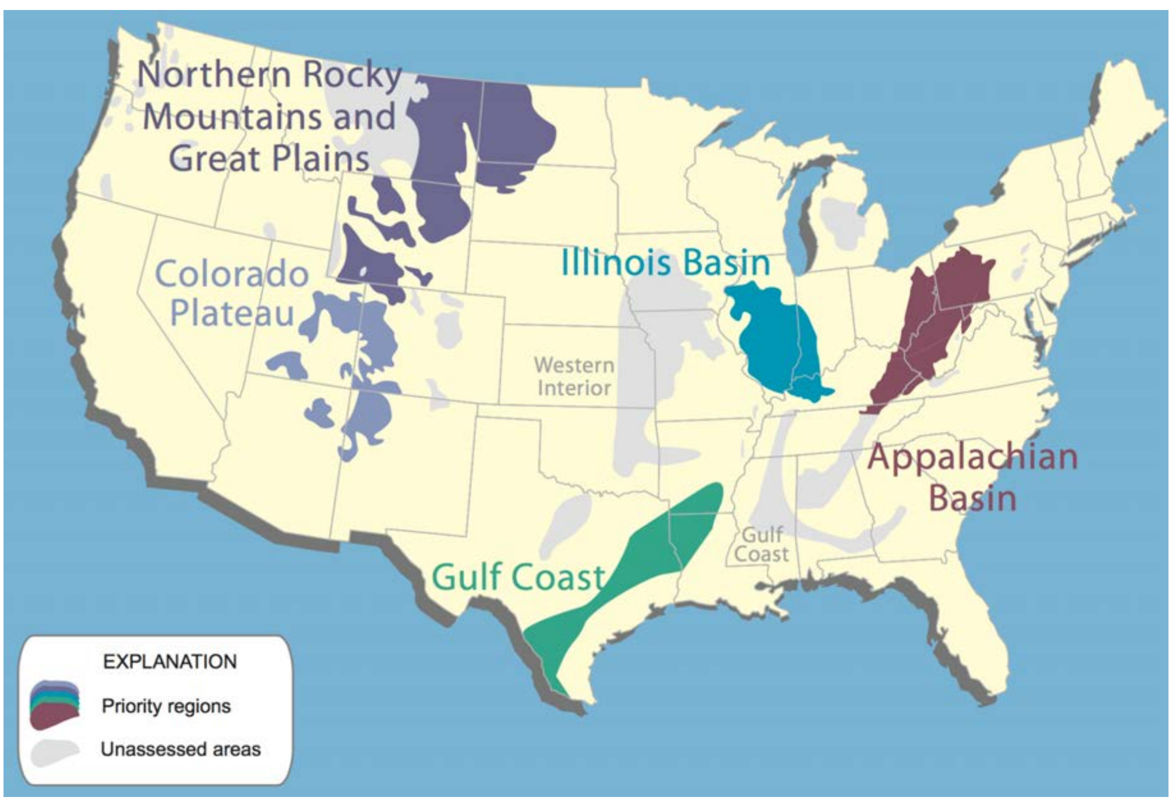
Notes: Reproduced from Energy Information Administration (2015).

Figure A.4: U.S. bituminous coal production and mechanization over time



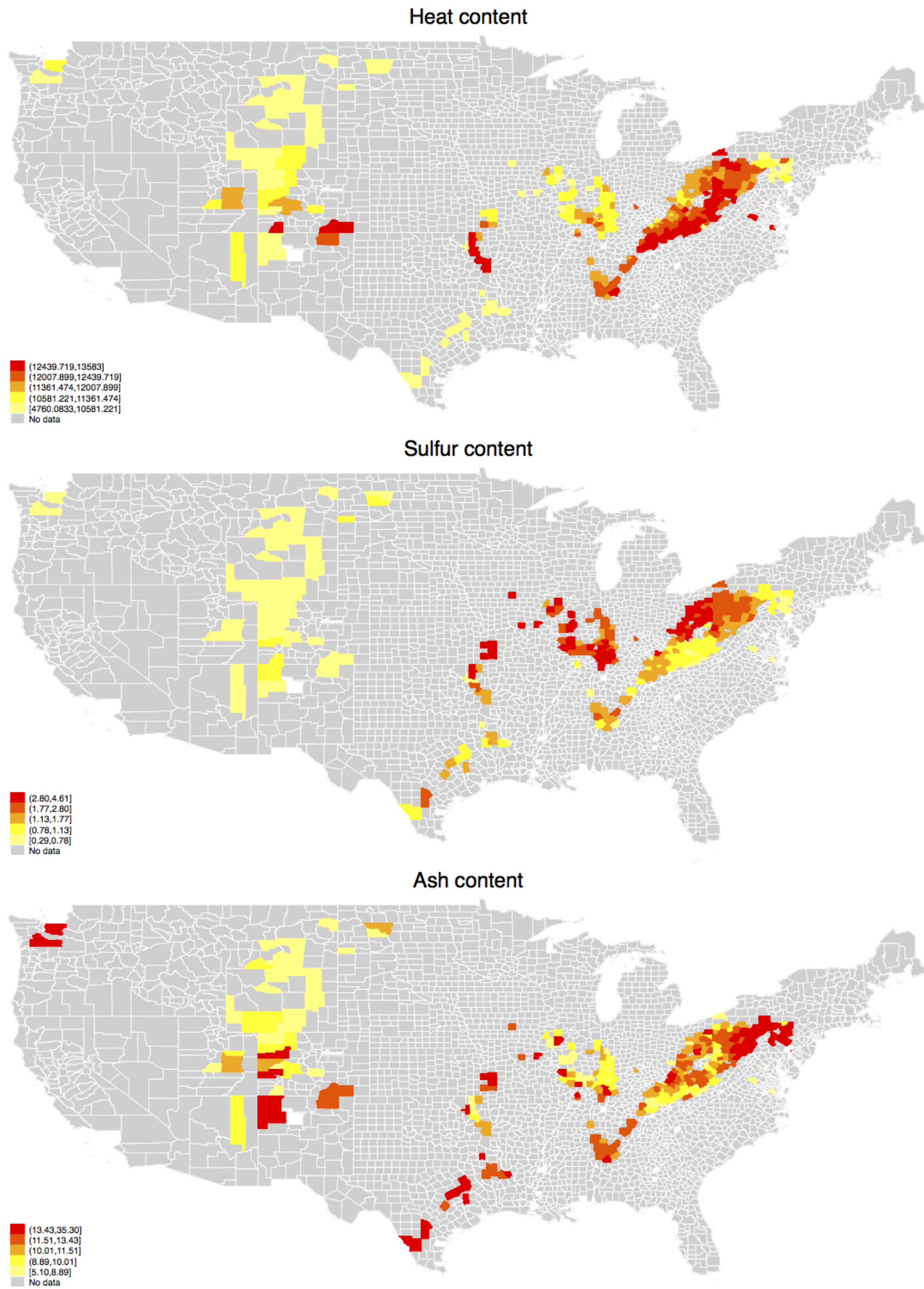
Notes: Bituminous coal production overall (blue) and from mechanized extraction (red). Source: U.S. Census Bureau (1975).

Figure A.5: U.S. coal basins



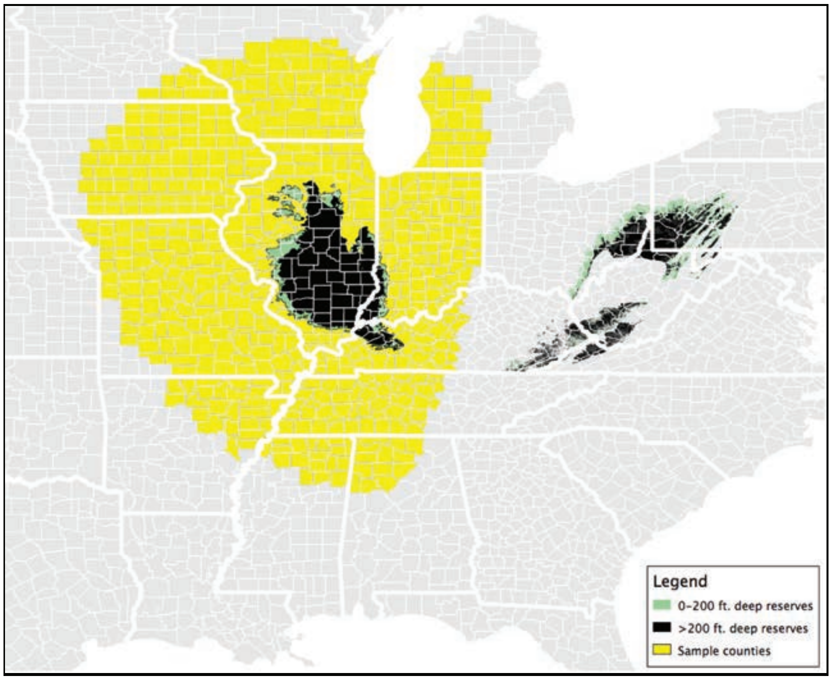
Notes: U.S. coal basins. Reproduced East (2012).

Figure A.6: Coal quality across the U.S.



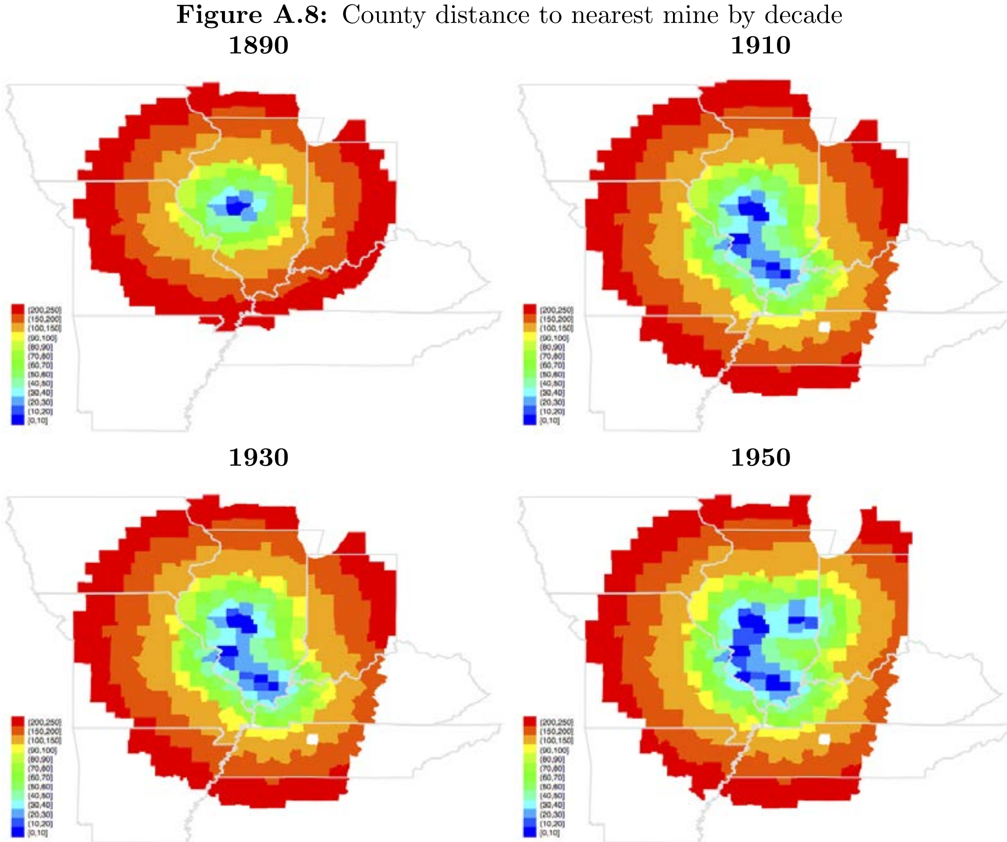
Notes: Maps show county-level average heat (in BTUs per lb, top panel), sulfur (in % of weight, middle panel), and ash (in % of weight, bottom panel) content of coal produced during 1990-1999. Source: EIA 423 forms.

Figure A.7: Sample counties and coal resources



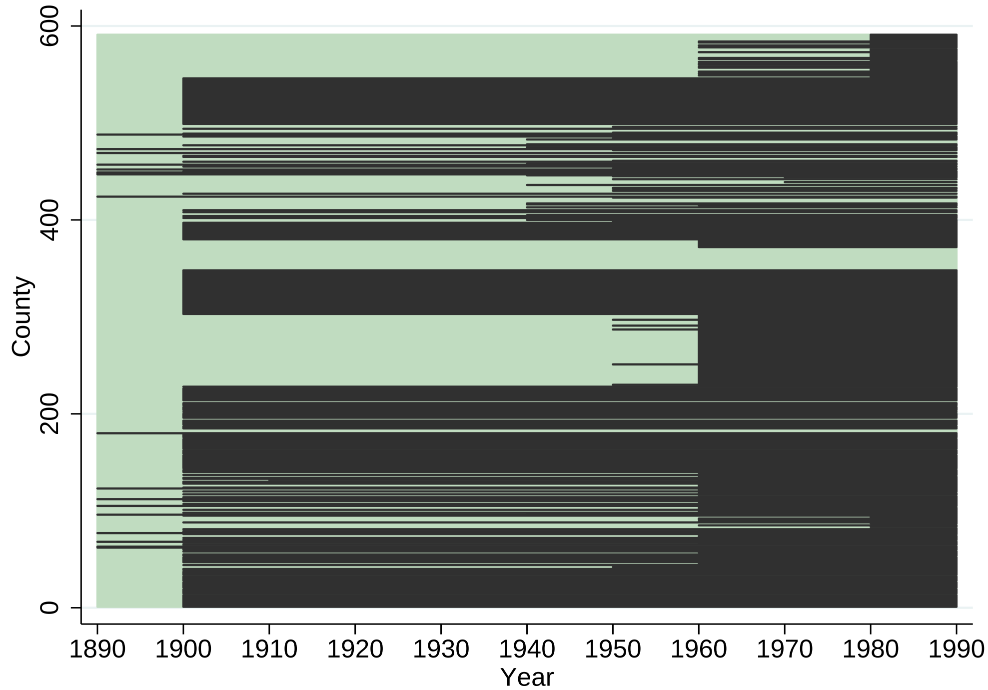
Notes: Counties are included in the sample (in yellow) if area-weighted county centroid is < 250 miles from nearest Illinois coal and closer to nearest Illinois coal than to nearest Appalachian coal. Illinois and Appalachian coal by depth also shown.

Figure A.8: County distance to nearest mine by decade



Notes: County distance to nearest mine from 1890-1950 over sample counties. Blank areas show counties that are more than 250 miles from nearest mine. See Section 4.1 for details.

Figure A.9: Timing of obsolescence for each sample county

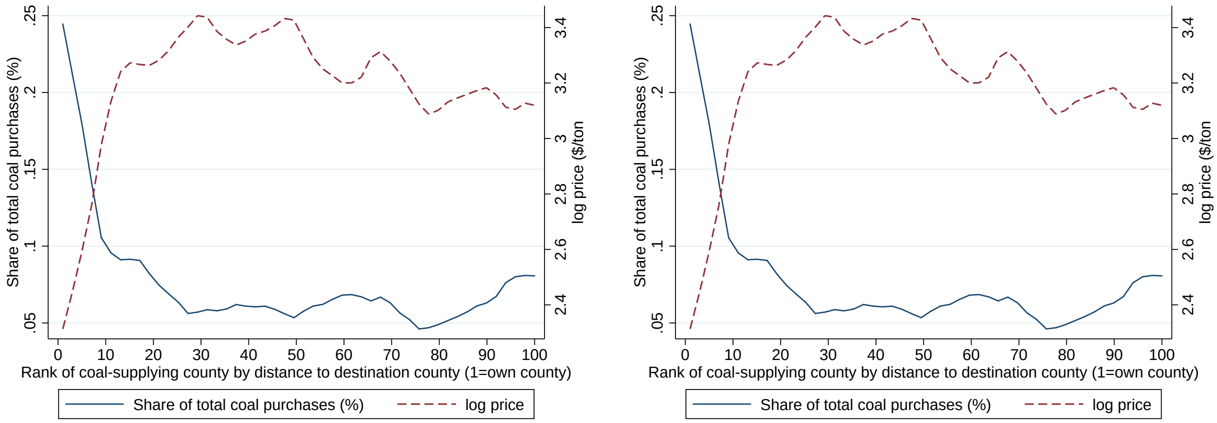


Notes: Plot shows event treatment timing for counties within the sample area (see Figure A.7). Light shading indicates period when nearest mine was shallow. Dark shading indicates period after a closer deep mine opens.

Figure A.10: Share of delivered coal price due to transport costs at the national level

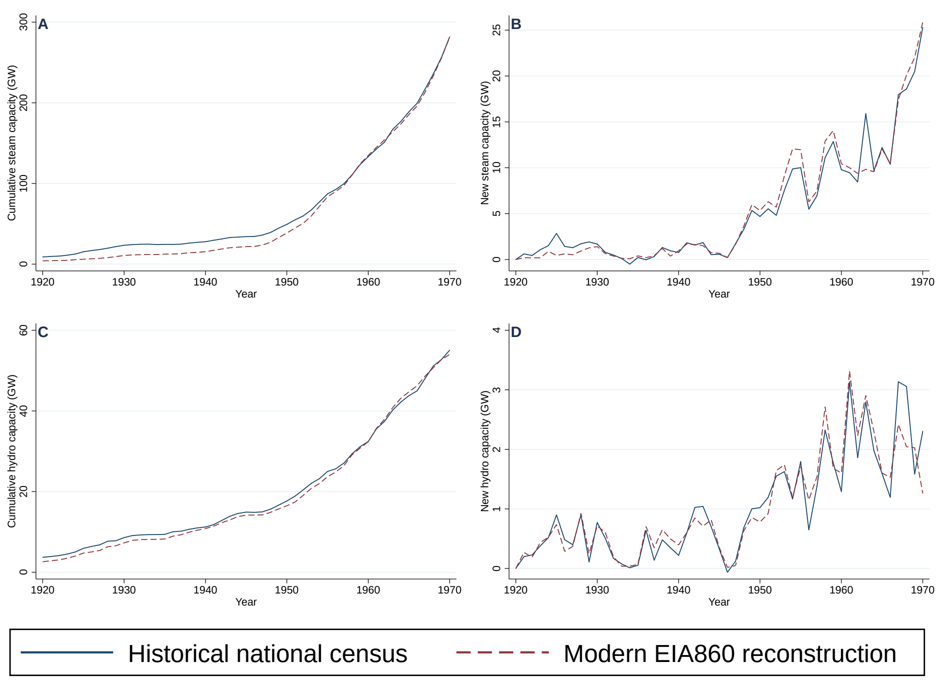
Notes: Historical national share of delivered coal price due to transport costs (1902-2007). Source: McNerney, Farmer and Trancik (2011)

Figure A.11: Testing delivered coal price definition from Hotelling's location model



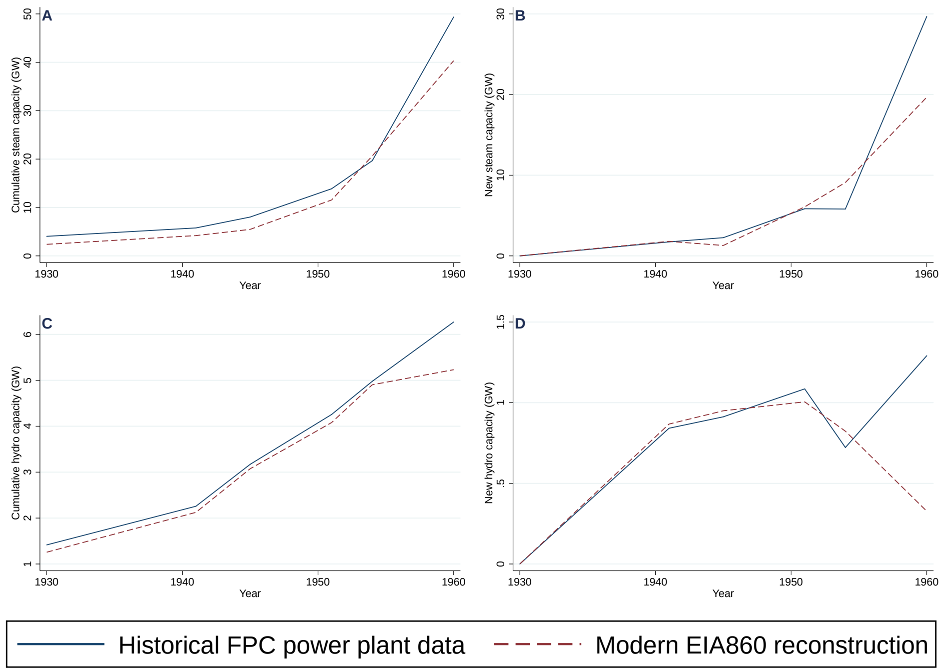
Notes: Share of total coal purchase and log delivered coal price for destination county by rank accordingly to distance. Data averaged over 1990-1999. Right panel includes all U.S. counties. Left panel includes counties in Alabama, Arkansas, Iowa, Illinois, Indiana, Kentucky, Minnesota, Missouri, Mississippi, Tennessee, and Wisconsin. Source: EIA 423 forms.

Figure A.12: Comparing historical census and EIA 860-constructed capacity nationally



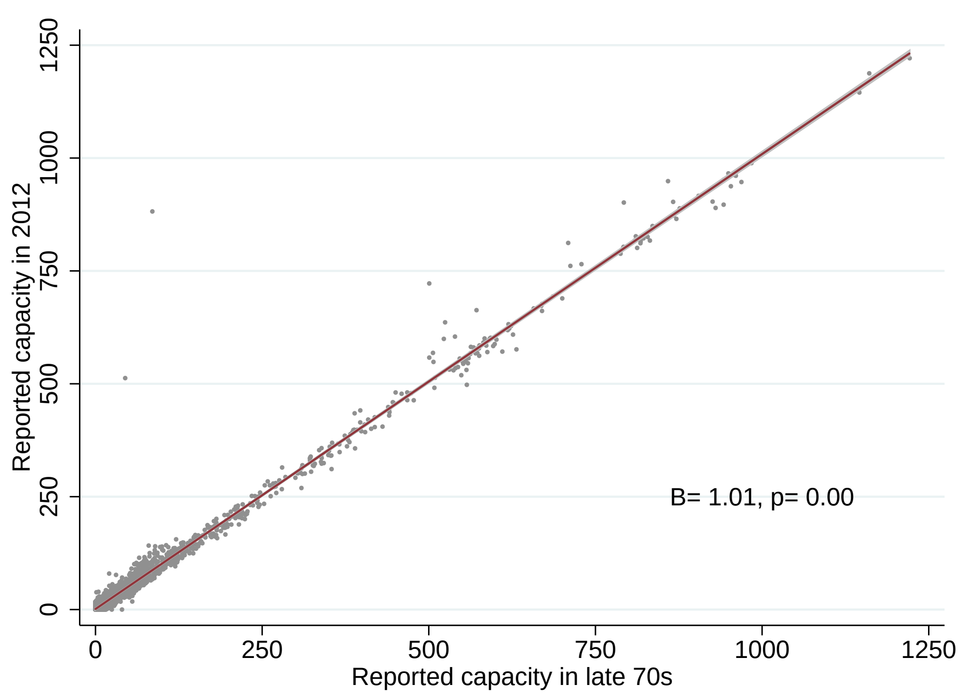
Notes: Panel (A) compares aggregate U.S. steam cumulative capacity from 1920-1970 from the U.S. historical census (U.S. Census Bureau, 1975) (solid blue) against values constructed from modern EIA 860 records (dashed red). Panel (B) compares new capacity. Panels (C) and (D) the same as Panels (A) and (B) respectively for aggregate U.S. hydro capacity.

Figure A.13: Comparing historical census and EIA 860-constructed capacity for sample states



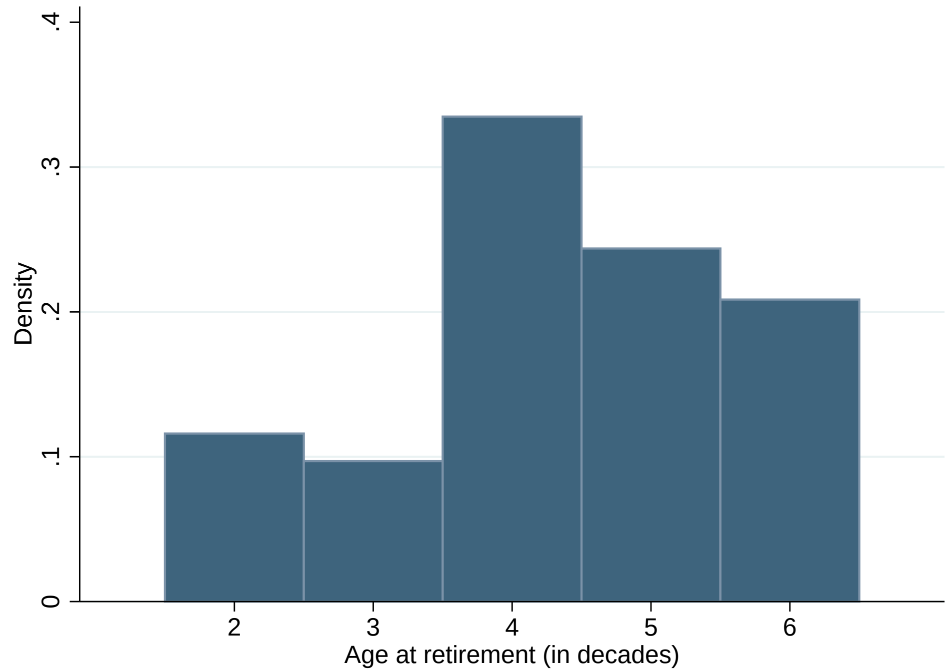
Notes: Panel (A) compares aggregate steam cumulative capacity from 1930-1960 within estimating sample states using power-plant level data from historical FPC data (see Lewis (2014)) (solid blue) against values constructed from modern EIA 860 records (dashed red). Panel (B) compares new capacity. Panels (C) and (D) the same as Panels (A) and (B) respectively for aggregate hydro capacity within estimating sample states. States include Alabama, Arkansas, Iowa, Illinois, Indiana, Kentucky, Minnesota, Missouri, Mississippi, Tennessee, and Wisconsin.

Figure A.14: Comparing generator-level capacity in late 1970s and 2012 EIA forms



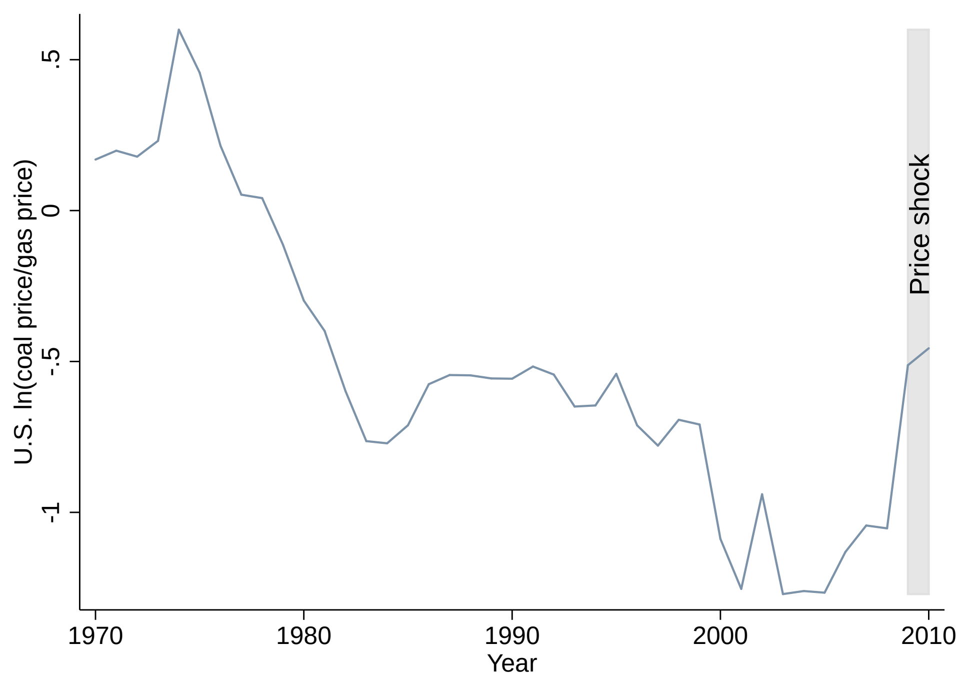
Notes: Scatter shows the reported generator capacity in the 2012 EIA 860 forms against the reported generator capacity in the late 1970's in the "Inventory of Power Plants in the United States"

Figure A.15: Lifespan distribution of retired generators



Notes: Distribution of lifespan for retired generators between 2 and 6 decades old. Generators located in Alabama, Arkansas, Iowa, Illinois, Indiana, Kentucky, Minnesota, Missouri, Mississippi, Tennessee, and Wisconsin. Source: EIA 860 forms.

Figure A.16: U.S. coal relative to natural gas sales prices



Notes: Log ratio of U.S. sales coal price over natural gas price. Source: Energy Information Administration (2012).

Appendix Tables

Table A.1: Checking observed county-level and constructed delivered coal price

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome is log delivered price of coal					
log dist. to nearest mine	0.38***	0.38**	0.55***	0.58***	0.33***	0.34**
	[0.14]	[0.14]	[0.16]	[0.16]	[0.11]	[0.11]
No. of counties	377	153	392	153	353	133
Decade	1970s	1970s	1980s	1980s	1990s	1990s
Sample	All	Sample counties	All	Sample counties	All	Sample counties

Notes: Each column is a county-level cross-sectional regression of delivered coal price from EIA 423 forms on distance to nearest mine. All models include state fixed effects. Columns (1) and (2) use 1970s values. Columns (3) and (4) use 1980s values. Columns (5) and (6) use 1990s values. Columns (1), (3), and (5) includes all U.S. counties with coal-fired power plants. Columns (2), (4), and (6) includes only counties in Alabama, Arkansas, Iowa, Illinois, Indiana, Kentucky, Minnesota, Missouri, Mississippi, Tennessee, and Wisconsin. Robust standard errors clustered at the county level. *** p<0.01, ** p<0.05, * p<0.1

Table A.2: Data consistency across EIA-860 forms (1990-2012)

	Percent of generators with different reported values			
Number of different values	Primary fuel	Capacity	Opening year	Retirement year
0	94.25	74.78	96.88	80.01
1	1.49	2.62	0.52	1.98
2	0.57	1.81	0.44	2.92
3	0.23	1.07	0.08	0.87
4	0.2	1.26	0.08	0.87
5	0.51	0.85	0.19	1.5
6	0.46	0.69	0.24	0.67
7	0.29	0.95	0.29	0.72
8	0.27	0.53	0.22	1.08
9	0.2	0.79	0.14	1.74
10	0.25	0.72	0.24	1.04

Notes: Row indicates the number of values from 1990-2011 EIA-860 forms that was different from the 2012 EIA-860 form. Column shows generator-level variables. Each cell shows the percentage of EIA-860 forms from 1990-2011 with a reported value that is different from that reported in 2012. For example, Row 1, Column 1 indicates that 94.25% of generators reported the same primary fuel from 1990-2011 as was reported in 2012.

Table A.3: Data consistency in reported primary fuel between late 1970s and 2012 EIA-860

Primary fuel in 1970s	Primary fuel in 2012				
	Coal	Hydro	Nat. gas	Nuclear	Oil
Coal	92.2	0.0	5.9	0.0	1.1
Hydro	0.0	100.0	0.0	0.0	0.0
Nat. gas	0.8	0.0	77.2	0.0	21.9
Nuclear	0.0	0.0	0.0	100.0	0.0
Oil	1.0	0.0	24.4	0.0	74.6

Notes: Each row shows the distribution of reported primary fuel in the 2012 EIA-860 forms conditional on the primary fuel reported in the 1970s from (Energy Information Administration, 1980). For example, 92.2% of generators which reported to use coal in the 1970s also reported to use coal in 2012.