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, Revised September 2018

I thank Dan Benjamin, Javier Birchenall, Hoyt Bleakley, Chris Costello, Melissa Dell, Jonathan Dingel, Walker Hanlon, Kelsey Jack, 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 many seminar participants for helpful comments. I am grateful to Sarah Anderson, Kira Fabrizio, Jessika Trancik, and Catherine Wolfram for sharing data. Kayleigh Campbell Bierman, Andy Bilich, Danae Hernandez Cortes, and Michael Pizzi provided excellent research assistance. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2016 by Kyle C. Meng. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Estimating Path Dependence in Energy Transitions
Kyle C. Meng
NBER Working Paper No. 22536
August 2016, Revised September 2018
JEL No. N51,N52,O41,Q35,Q43,Q54,Q58

ABSTRACT

How can an economy transition from dirty to clean inputs? Structural change models posit that when an economy exhibits strong path dependence in its transitional dynamics, a temporary price shock can trigger permanent change in the composition of inputs. This paper examines whether the U.S. electricity sector exhibits such dynamics across the 20th century in its use of coal, the fossil fuel that contributes most to climate change. Exploiting shocks to local coal prices driven by technological advances in coal mining, I find increasing imbalance in the coal composition of electricity capital lasting ten decades following a shock. Additional tests detect increasing returns to scale as the underlying mechanism. I develop a model of scale-driven structural change which links my reduced-form estimates to a formal definition of strong path dependence and to a key structural parameter found in a broad class of structural change models. Further simulations characterize the conditions under which a temporary climate policy can trigger a sustained future energy transition away from coal.

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

An economy tends to consume more of its abundant resources, turning to alternatives only after supply of that resource has sufficiently depleted (Herfindahl, 1967). However, when that resource is associated with substantial social costs, an earlier switch may be desirable. Coal is a textbook example of such a resource. It is the most abundant fossil fuel for producing electricity, representing over half of energy contained in global fossil fuel deposits today (BP, 2017). Coal is also dirty. Local pollution from coal provided some of the earliest examples of external costs in economics (Pigou, 1920; Coase, 1960), which has since been studied extensively.[1] Globally, coal is responsible for over half of the anthropogenic carbon dioxide emitted since 1750 (Boden and Andres, 2013). As a consequence, it is widely recognized that any strategy for addressing climate change must shift electricity production around the world away from coal.

How can such a transition be induced? Economics offers two perspectives. In the traditional view, the composition of resources consumed in an economy is primarily determined by relative supply. Because coal is more abundant and thus cheaper than other fuels, a long-term transition away from coal requires a *permanent* policy intervention that raises its price. If this policy were ever removed, coal consumption and carbon emissions would eventually resume their previous upward trajectories. This behavior appears in economies that exhibit either no or weak path dependence in energy transitions. Unfortunately, permanent policy interventions are unrealistic when governments cannot commit to long-term policies. Indeed, the short history of climate policies to date is filled with examples of policy revisions, reversals, and withdrawals.[2]

In contrast, recent models of structural change posit that in the presence of sufficiently strong transitional dynamics, a large but *temporary* intervention that exogenously increases the price of coal can overcome the forces of its abundant supply. In such circumstances, a sustained long-term transition towards cleaner fuels could be achieved even after the intervention is lifted (Acemoglu et al., 2012, 2016). Economies with this feature are broadly characterized as having strong path dependence in energy transitions.[3] While strong path

[1] Recent empirical studies have shown that local pollution from coal burning leads to increased health risks and reduced property values (Chay and Greenstone, 2003, 2005; Barreca, Clay and Tarr, 2014; Clay, Lewis and Severnini, 2016; Beach and Hanlon, 2018; Hanlon, 2016).

[2] For example, the U.S. recently withdrew from the U.N. Paris Agreement and announced plans to review the Clean Power Plan, its domestic climate policy. In 2011, Canada withdrew from the Kyoto Protocol, the preceding U.N. climate agreement, several years after the Protocol entered into force. Key details of the E.U. and California carbon trading markets have been revised since their inceptions. See Acemoglu and Rafey (2018) for other examples.

[3] A related literature from the economics of technological change discusses how path dependence can differentially lock-in certain technologies over time (Schumpeter, 1942; Schmookler, 1966; David, 1985; Aghion

dependence can arise from various mechanisms, in a reduced-form sense, it occurs whenever an exogenous, temporary shock to fuel prices triggers an increasing fuel imbalance in subsequent periods. Whether such dynamics actually exist in the electricity sector, however, is an open empirical question.

This paper estimates path dependence in historical U.S. energy transitions over the 20th century to shed light on how coal has come to dominate the U.S. electricity sector. In doing so, this paper also provides new insights on the conditions for triggering sustained future energy transitions away from coal and towards lower carbon emissions. Exploiting shocks to local coal prices from the changing accessibility of subsurface coal, I find reduced-form evidence consistent with strong path dependence in the transitional dynamics of the U.S. electricity sector. To interpret these results, I develop a model of structural change for the electricity sector which allows for a formal definition of strong path dependence, a mapping between my reduced-form estimates and a key structural parameter, and simulations of future energy transitions away from coal.

There are several empirical challenges to estimating path dependence in energy transitions. First, to be relevant for future long-term transitions, one needs a data setting that can capture changes in the fuel composition of electricity capital. Because power plants last multiple decades, this requires having local coal prices going back to the start of the 20th century. Unfortunately, an extensive search of historical records revealed such data were either never historically collected or no longer exists today. Second, for causal inference, identifying variation in coal prices must not only be uncorrelated with unobserved contemporaneous determinants of electricity capital, but must also be uncorrelated with such determinants in subsequent periods. This latter requirement ensures that lagged effects can be interpreted as path dependence following a transitory shock, as distinguished from the effects of some persistent determinant (see Bleakley and Lin (2012)).

To meet these challenges, I construct shocks to local coal prices following the introduction of mechanized coal mining around the early 20th century. Mechanized mining allowed extraction over previously inaccessible coal held in deep underground deposits. Its introduction, together with the location and depth of coal resources, altered the transport distance between counties and their nearest coal mine, inducing some counties to switch their coal supply from shallow to deep coal mines. These transport distances act as supply-side shocks which altered the spatial distribution of delivered coal prices. Because my distance measure is constructed by interacting the spatial structure of subsurface coal geology with the aggregate introduction of mechanized mining, it is less likely to be correlated with unobserved contemporaneous and lagged determinants of electricity capital. Furthermore, construction

and Howitt, 1992; Arthur, 1994).

of these shocks over the 20th century requires only the location, depth, and operating years of coal mines, data which was recently made available from geological archives for the Midwestern U.S.

I examine the consequences of changing transport distance in an event study setting using county-by-decade panel data. Because the initial switch from shallow to deep coal occurs at different moments in time for different sample counties, my event study specification allows for the inclusion of state-by-decade and county fixed effects. I find that distance to the nearest shallow mine prior to the switching event has increasing influence on the fuel composition of electricity capital, with lagged effects detected up to ten decades later. Notably, the time-pattern of lagged effects displays discrete jumps at two and seven decades after the event. These jumps correspond to the expected timing of when two subsequent vintages of electricity capital would be constructed and suggest that my estimates are capturing effects on new capital investments. I do not find lead effects. Results from additional robustness checks address other identification concerns arising from local electricity demand shocks, coal mine heterogeneity and market power, as well as a host of potential concerns related to data construction and statistical modeling.

Path dependence in energy transitions emerges from a combination of "push" and "pull" forces, each of which may arise from various mechanisms. Prior literature suggests two channels that may amplify transitions in the electricity sector: 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 mechanisms, I develop a model of electricity production that includes vintages of electricity capital with fixed factor productivities (Komiya, 1962; Joskow, 1985, 1987; Atkeson and Kehoe, 1999) and increasing returns to scale across capital vintages. To provide a dampening force against runaway transitions, my model further includes imperfect substitutability between electricity from coal and other fuels (Acemoglu et al., 2012). I find empirical support for the presence of scale over productivity accumulation effects as the relevant mechanism behind my reduced-form estimates of path dependence. Additional tests fail to detect other mechanisms related to cost-of-service electricity regulation, the U.S. Clean Air Act, coal procurement contracts, increasing returns in coal transportation, and residential household sorting.

Having isolated the relevant mechanism, I then turn to a structural interpretation of my reduced-form estimates. I first use my theoretical framework to formally define strong path dependence in energy transitions, which occurs whenever the push from increasing returns to scale offsets the pull from the imperfect substitutability between electricity from coal and other fuels. This formal definition enables a mapping between my reduced-form estimates and the elasticity of substitution, a key parameter that appears across a broad class of structural change models driven by supply-side forces, first posited by Baumol (1967).[4] As such, recovery of this parameter allows my findings to speak to other energy transitions where the underlying mechanism may differ from this study. For example, in the context of Acemoglu et al. (2012)'s model of optimal climate policy under directed technical change, my recovered elasticity of 4.5 implies that a temporary policy intervention would be sufficient to avoid a future where the climate deteriorates to a point beyond recovery.

Finally, I combine my reduced-form results and theoretical framework to draw lessons for future clean energy transitions. Strong path dependence implies it is possible for a temporary fuel price intervention to induce permanent fuel switching. But under what conditions? The magnitude and/or duration of the required intervention depends on baseline relative coal prices. If baseline relative coal prices are low, as they are in the U.S., a very large and/or long-lasting intervention is needed to prevent the same forces of path dependence from once again favoring coal after the intervention. To analyze what kind of interventions are needed, I simulate future U.S. electricity sector carbon emissions triggered by temporary relative coal price shocks of varying magnitude and duration. To ground simulations in recent developments, the magnitude of shocks are benchmarked to recent high relative coal prices following the introduction of natural gas hydraulic fracturing. Simulations show that for a better than 50% chance of achieving a permanent switch away from coal and thus weakly declining carbon emissions, one needs recent relative coal prices to last at least three decades. If only a one decade intervention is considered, the relative coal price shock must then be three times higher.

Two related papers examine how fuel price shocks at higher temporal frequencies affect the composition of inputs in the energy sector. Aghion et al. (2016) explore firm-level path dependence in clean and dirty patenting activity for the automobile sector across 80 countries in response to annual oil prices during 1986-2005. Knittel, Metaxoglou and Trindade (2015) examine U.S. electricity fuel-switching due to monthly variation in natural gas prices during 2003-2012. My study has the advantage of exploring transitional dynamics over a longer time horizon in order to detect capital changes. Knowledge of long-run effects is particularly important given that sufficient future reductions in carbon emissions is unlikely to occur with the existing stock of electricity capital.

The identification strategy used in this paper relates to a growing empirical literature that examines dynamic effects following some change in the productivity of a natural resource. This literature is generally classified by whether a new technology "turns on" or "turns off"

[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).

the productive use of some resource. The adjustment literature estimates lagged effects after a resource becomes productive (Feyrer, 2009; Nunn and Qian, 2011; Hornbeck and Keskin, 2014). The path dependence literature, which this paper belongs to, explores the lagged effects of a previously productive or accessible resource (Bleakley and Lin, 2012; Dell, 2012; Hornbeck, 2012; Severnini, 2014; Hornbeck and Naidu, 2014; Glaeser, Kerr and Kerr, 2015). Finally, the use of transport distance as identifying variation can also be found in papers by Feyrer (2009), Atkin and Donaldson (2015), and Donaldson (2018).

The remainder of the paper is organized as follows: Section 2 presents motivating evidence on the U.S. electricity sector. Section 3 details statistical challenges and proposes a solution. Section 4 discusses data construction and verification checks. Section 5 presents reduced-form evidence of path dependence and related 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 a structural parameter, and simulate future carbon emissions. Section 8 concludes.

2 Background: Why is the U.S. so dependent on coal?

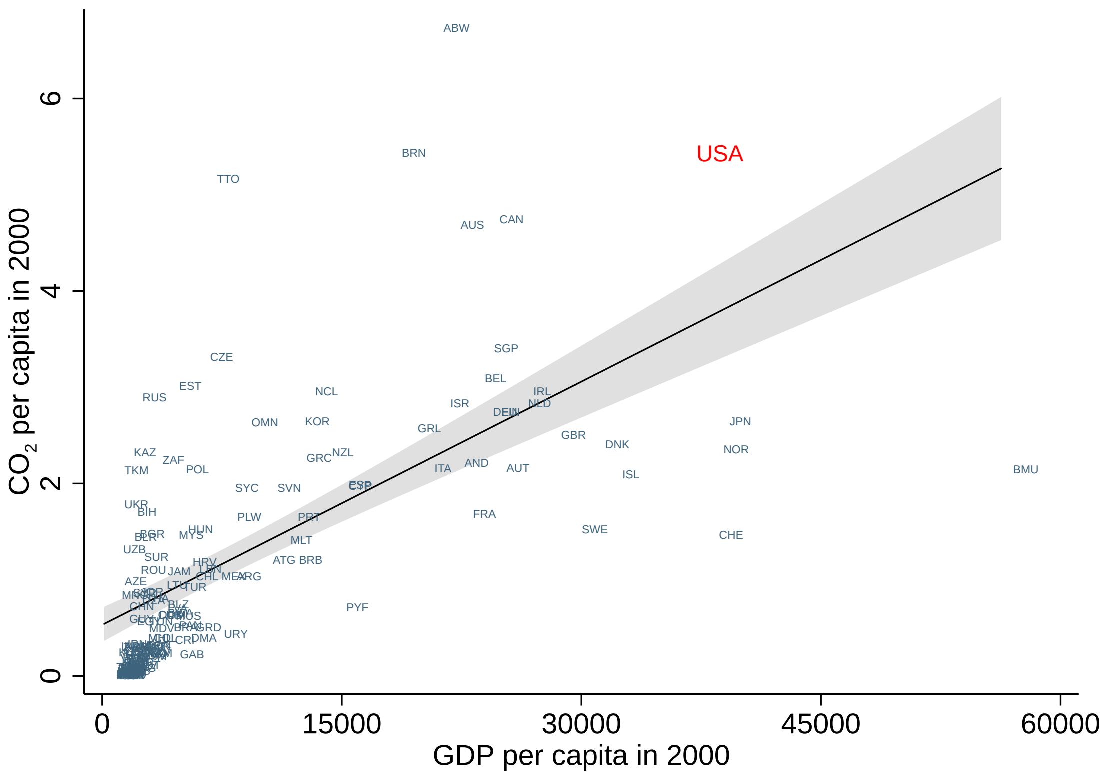
The United States has one of the most carbon-intensive economies in the world. Figure 1 shows carbon dioxide (CO_2) emissions per capita plotted against GDP per capita for non-OPEC countries in 2000. U.S. emissions per capita is nearly 2 standard deviations higher than what income alone would predict. This reflects the U.S. electricity sector's heavy reliance on coal, the most carbon-intensive of energy inputs.[5] Since the 1960s, roughly 40% of U.S. electricity production has been based on coal (Energy Information Administration, 2012). Why is the U.S. electricity sector so dependent on coal? Many observers point to the country's world-leading coal resources. A casual examination of historical local coal use, however, suggests that coal abundance provides only a partial explanation.

The typical measure of capital size in the electricity sector is an electric generator's capacity, or the maximum amount of electricity that a generator can produce. Generators, in turn, are linked to boilers that consume a particular fuel, to form a generating unit. Thus, a natural measure of the coal composition of electricity capital is the ratio of capacity from coal-fired generating units to the capacity of generating units using other fuels. I call this measure relative coal capital.[6]

[5] Bituminous coal, the most common type of coal for electricity, produces 206 lbs of CO_2 per million British Thermal Units (BTU). By contrast, oil and natural gas produces 157 and 117 lbs of CO_2 per million BTU, respectively.

[6] Relative coal capital is used throughout this paper in large part because many predictions from standard models of structural change (discussed further in Sections 6 and 7) are expressed in ratios of factor inputs

Figure 1: CO_2 emissions intensity and income in 2000



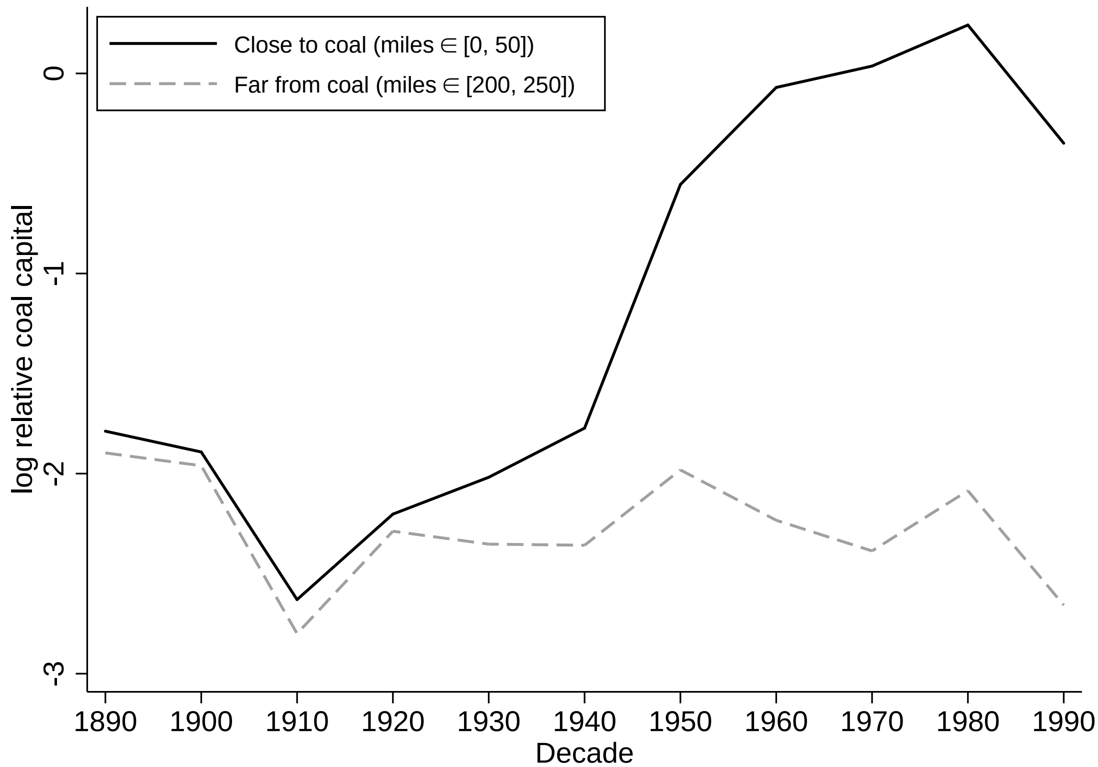
NOTES: Plot shows carbon dioxide emissions (in tons of carbon) per capita against GDP (in nominal USD) per capita in 2000. Linear regression fit shown with 95% confidence interval. OPEC countries excluded. Data from Boden and Andres (2013) and World Bank (2014).

Figure 2 examines U.S. relative coal capital at the county-by-decade level (see Section 4.2 for data construction details).[7] It plots average log relative coal capital over 1890-1990 separately for counties that are close to (between 0 and 50 miles) and further away (between 200 and 250 miles) from coal resources in the Illinois Coal Basin, the basin which is the focus of this paper (for reasons discussed in Section 4.1). Consistent with a coal supply argument, counties closer to coal resources consistently have higher relative coal capital throughout the 20th century. However, the gap between these two sets of counties grows dramatically as the 20th century progresses. Such a pattern cannot be explained by coal supply alone. If anything, that argument predicts convergence, not divergence, in relative coal capital as locations closer to coal deplete their supplies more rapidly over time. Instead, Figure 2 hints at the presence of path dependence, whereby past relative coal capital somehow affects its future values. Causal detection of path dependence, however, requires additional statistical assumptions, which I detail next.

across sectors.

[7] Because electricity capital last multiple decades, I use a decade as the time-step for all empirical analyses.

Figure 2: Relative coal capital for counties close to and far from Illinois Coal Basin



NOTES: Plot shows log relative coal capital averaged across counties with centroids that are less than 50 miles (solid black) and between 200 and 250 miles (dashed gray) from the nearest coal resource in the Illinois Coal Basin for each decade from 1890 to 1990.

3 Empirical framework

This section begins by describing the empirical challenges to identifying path dependence in the coal composition of electricity capital. I then discuss how the introduction of mechanized mining offers an opportunity to overcome these challenges. To focus for now on data and statistical issues, this section defines path dependence and its strength within a reduced-form setting. Section 7 will introduce a model that provides formal definitions.

3.1 Challenges to identifying path dependence

Statistically, path dependence is present whenever an *exogenous* and *temporary* shock continues to exert influence on an outcome of interest after the shock. Path dependence in structural change occurs when that shock subsequently alters a sector's composition of inputs. To illustrate the challenges with estimating path dependence in energy transitions, I begin with a simple empirical framework. There are two decades, $t \in \{1, 2\}$, and two fuel-specific intermediate sectors of electricity production, $j \in \{c, n\}$. Sector c produces electricity using coal; sector n produces electricity using other fuels. The outcome of interest is relative coal capital in county i, $\widetilde{K}_{it} = \frac{K_{cit}}{K_{nit}}$. w_{cit} is the delivered coal price. Demand for relative coal capital in the second period takes the following form

$$\ln \widetilde{K}_{i2} = \beta \ln w_{ci1} + \pi \ln w_{ci2} + \xi_{i2} \tag{1}$$

where ξ_{i2} captures unobserved determinants of demand. Two statistical assumptions are needed for β to have a path dependence interpretation. The first is $E[w_{ci2}\xi_{i2}|w_{ci1}] = 0$, which states that contemporaneous coal prices are exogenous. This assumption would be violated if, for example, relative coal capital and coal prices were simultaneously determined. The second identifying assumption is $E[w_{ci1}\xi_2|w_{ci2}] = 0$. Specifically, it states that past coal prices are uncorrelated with contemporaneous determinants of relative coal capital. If this "exclusion restriction" assumption is satisfied, β will only capture the indirect effect of past price shocks on current relative capital as mediated through past relative capital. This assumption would be violated if, for example, an unobserved time-invariant determinant of relative capital such as a county's geography is correlated with past coal prices. In that case, variation in past prices has a direct influence on current relative capital. If so, β could not isolate path dependence following a temporary price shock from the effects of some persistent unobserved determinant. When both assumptions are satisfied, price shock effects that grow over time provide evidence for strong path dependence in energy transitions. Conversely, effects that eventually dissipate would suggest the presence of weak path dependence.

In practice, another complication arises when estimating equation (1), which implicitly assumes that the lagged effect β can be detected within a single decade. Electricity capital lasts multiple decades. To detect effects on subsequent new capital investments, one needs county-level coal prices across much of the 20th century. Unfortunately, to the best of my knowledge, such historical data were either never collected or, if collected, is no longer available today (see Appendix B for a summary of historical data collection and availability).

3.2 Solution: local coal supply shocks from mechanized mining

To address these three empirical challenges, I employ a proxy measure for county-level delivered coal prices that require weaker identifying assumptions and span a 110-year period. The basic idea is to construct local coal supply shocks driven by the interaction between the time-invariant spatial structure of subsurface coal and the timing of when mechanized extraction allowed for deep coal mining.

Mechanized mining and access to deep coal Prior to the 20th century, most coal in the U.S. was manually extracted which limited mining to resources generally less than 200 feet from the surface (Fisher, 1910; Speight, 1994). Mechanized mining was introduced around the turn of the century and eventually came to dominate coal extraction. The expansion of mechanized mining is shown in Figure A.1. Between 1890-1930, nearly the entire production increase in bituminous coal - the variety most used for electricity - came from mechanized extraction (U.S. Census Bureau, 1975). The main benefit of mechanization was the introduction of mechanized drills that allowed for the excavation of previously inaccessible deep coal resources. Access to this newly extractable coal resource altered the spatial distribution of delivered coal prices.

Constructing distance-based local coal supply shocks To illustrate how mining technology interacts with the location and depth of underground coal, I expand the previous framework to include coal resources and mines. Let $\ell = 1, \dots, L$ index locations that overlie coal resources. The depth of coal in these locations is represented by the vector $\mathbf{z} = (z_1, ...z_L)$ where z_ℓ is the depth of coal in location ℓ. Whether or not coal in a given location is mined depends on the technology available in period t, $\chi_t \in \{0, 1\}$. In the first period, $\chi_1 = 0$ and only locations with shallow coal, $z_\ell < 200$ft, can be mined. In the second period, $\chi_2 = 1$ and all locations with coal can be mined. In this setting, the set of active mines in each period, denoted as M_t, is driven only by the time-invariant cross-sectional vector of coal depth, $\mathbf{z}$, and time-varying mining technology, χ_t. Let $m_t \in M_t$ index a mine operating in decade t.

To construct a local supply shock to delivered coal prices, I enlist the Herfindahl Principle (Herfindahl, 1967), a common result from models of spatial competition. The Herfindahl Principle states that when a homogeneous resource is costly to transport across space, and homogeneous suppliers are perfectly competitive, a consumer buys from the physically nearest supplier at a price that is set, in part, by distance to that supplier.[8] For county i in decade t, distance to the nearest mine is

$$d_{it} = \min_{m_t \in M_t} \{||i - m_t||\}$$

where $||i - m_t||$ is the Euclidean distance between county i and coal mine m_t. Delivered coal

[8] More specifically, the Herfindahl Principle states that under perfect competition, a consumer will buy resources from the producer with the lowest cost of supplying to that consumer. When producers have homogeneous resource endowments and extraction costs, and transport costs scale with distance, the lowest cost producer for a consumer is its physically nearest producer. See Gaudet, Moreaux and Salant (2001) for a generalized setting with multiple, spatially differentiated, consumers. Robustness checks in Section 5.2 considers potential complications that arise when coal resources have heterogeneous quality and when coal mines have market power.

price can then be decomposed into

$$\ln w_{cit} = \ln d_{it} + \zeta_{it} \tag{2}$$

where ζ_{it} includes other supply-side factors. Inserting equation (2) into equation (1) yields

$$\ln \widetilde{K}_{i2} = \beta \ln d_{i1} + \pi \ln d_{i2} + \mu_{i2} \tag{3}$$

where $\mu_{i2} = \beta\zeta_{i1} + \pi\zeta_{i2} + \xi_{i2}$. $\ln d_{i2}$ is log distance to the nearest mine when mechanized mining is available. $\ln d_{i1}$ is log distance to the nearest mine before mechanized mining.[9] To ensure β has a path dependence interpretation, the updated exogeneity and exclusion restriction assumptions for equation (3) are $E[\ln d_{i2}\mu_{i2}|\ln d_{i1}] = 0$ and $E[\ln d_{i1}\mu_{i2}|\ln d_{i2}] = 0$, respectively.

Compared with directly using coal prices in equation (1), use of $\ln d_{it}$ has two distinct advantages. First, observe that only three elements go into constructing $\ln d_{it}$: (i) the time-invariant spatial distribution of coal resource depths, $\mathbf{z}$, (ii) the introduction of mechanized mining, χ_t, and (iii) the location of each county. Thus, construction of $\ln d_{it}$ requires historical data on the location, coal depth, and the operating years of coal mines, which unlike historical coal prices is available. Second, because $\ln d_{it}$ is constructed by interacting the spatial structure of subsurface coal geology with aggregate technological change, it is less likely to be correlated with unobserved contemporaneous and lagged determinants of relative coal capital, and thus more likely to identify path dependence. Additional identification concerns may still exist. I now discuss how my actual data setting addresses these concerns by allowing for a richer set of controls.

3.3 Regression specification

In practice, my data setting covers multiple counties observed over multiple decades. This allows for a generalized version of equation (3) offering four empirical advantages. First, I can include county fixed effects to control for unobserved time-invariant county characteristics. Second, because each county experiences the initial switch to buying from a deep coal mine at different moments in time, I can estimate an event study model where the timing of the event differs across counties. This feature allows me to estimate effects before and after the switching event that are separately identified from state-by-decade fixed effects. Third, the presence of multiple decades of data for each county allows estimation of lagged effects

[9] In this setting, it need not be the case that $\ln d_{i1} \neq \ln d_{i2}$ for all counties. The introduction of mechanized mining may not lead to the opening of a deep coal mine that is closer to a given county than an existing shallow coal mine.

over a sufficiently long time horizon that should capture new capital investments following the coal price shock. Finally, my model includes lead terms to examine whether there were differential pre-event trends in relative coal capital. Broadly speaking, all four elements address remaining concerns that local coal supply shocks driven by the interaction between the spatial distribution of coal resource depths and the introduction of mechanized mining may still be correlated with unobserved determinants of local coal demand and supply.

Denote $h = -2, \ldots, 10$ as the event-time index. For each county, there are three distinct periods, each occurring when the distance to its nearest coal mine changes. The focal event occurs when a county's nearest mine switches from a shallow to a deep coal mine for the first time. $h > 0$ denotes the period after that switch. $h = 0$ marks the period before the switch during which a county was closest to a shallow coal mine.[10] $\ln d_i^0$ is distance to that shallow coal mine and represents my local supply shock to delivered coal prices. $\ln d_i^0$ is a temporary shock in that once a county switches to deep coal, $\ln d_i^0$ should no longer have a direct effect on relative coal capital. Any lagged effects from $\ln d_i^0$ can therefore be interpreted as path dependence. $h < 0$ indicates an earlier period when a county's coal supplier was yet a different shallow coal mine. Using this event study setting setting, I estimate the following generalized version of equation (3) for county i in state s, during decade t

$$\ln \widetilde{K}_{it} = \sum_{\substack{-2\leq\tau\leq 10\\ \tau\neq 0}} \beta^\tau [\ln d_i^0 \times \mathbf{1}(\tau = h)] + \pi \ln d_{it} + \sum_{\substack{-2\leq\tau\leq 10\\ \tau\neq 0}} \gamma^\tau \mathbf{1}(\tau = h) + \rho_i + \phi_{st} + \mu_{ist} \quad (4)$$

where ρ_i and ϕ_{st} are county and state-by-decade fixed effects, respectively. γ^τ captures common event-time effects and μ_{ist} is an error term. π is the contemporaneous coal price effect. β^τ are my parameters of interest. When $\tau > 0$, β^τ are lagged effects of past coal price shocks and capture path dependence. When $\tau < 0$, β^τ are lead effects and test for the presence of differential pre-event trends. Equation (4) is my baseline specification. In Section 5.2, I also estimate variants of equation (4) to address other possible identification concerns arising from (i) electricity demand from the manufacturing sector, (ii) mine heterogeneity, and (iii) market power exerted by mines.

4 Data

This section first details how spatial data on coal resources and mines are used to construct my local coal supply shocks. I then describe how my main outcome variable, relative coal

[10] Note that $h = 0$ can span multiple decades if the shallow coal mine is a county's nearest supplier for some time.

capital, is constructed. For both variables, I present a series of tests to verify the assumptions behind my construction procedures.

4.1 Distance-based local coal supply shocks

The USGS National Coal Resource Assessment (NCRA) recently amassed and digitized spatial data on coal resources and mining that was previously separately held in the archives of various state geological agencies (East, 2012). Crucially, the NCRA provides GIS shape files for coal resources of different depths for each of the major U.S. coal basins, shown in Figure A.2. This paper focuses on counties located near the Illinois Coal Basin.[11] This spatial restriction is done for data availability and estimation reasons.

First, of the major basins assessed by the NCRA, there is data on the location, areal extent, and opening and closing years of coal mines only for the Illinois Basin.[12] This data goes back to 1890. The Illinois Basin is also advantageous for its geological properties. As Figure 3 shows, it has a dish-like shape with shallow coal resources in the outer regions and deeper resources near the center.[13] Coal across a single large deposit is less likely to be heterogeneous in quality. Indeed, data from the FERC-423 coal procurement forms displayed in Table A.1 show that among the five major coal basins, Illinois Basin coal has the second lowest heterogeneity in both heat and ash content. Other coal basins also exhibit more complicated subsurface structures with overlapping coal layers situated under the same location. In such cases, shallow coal mining could lead directly to deep coal extraction.

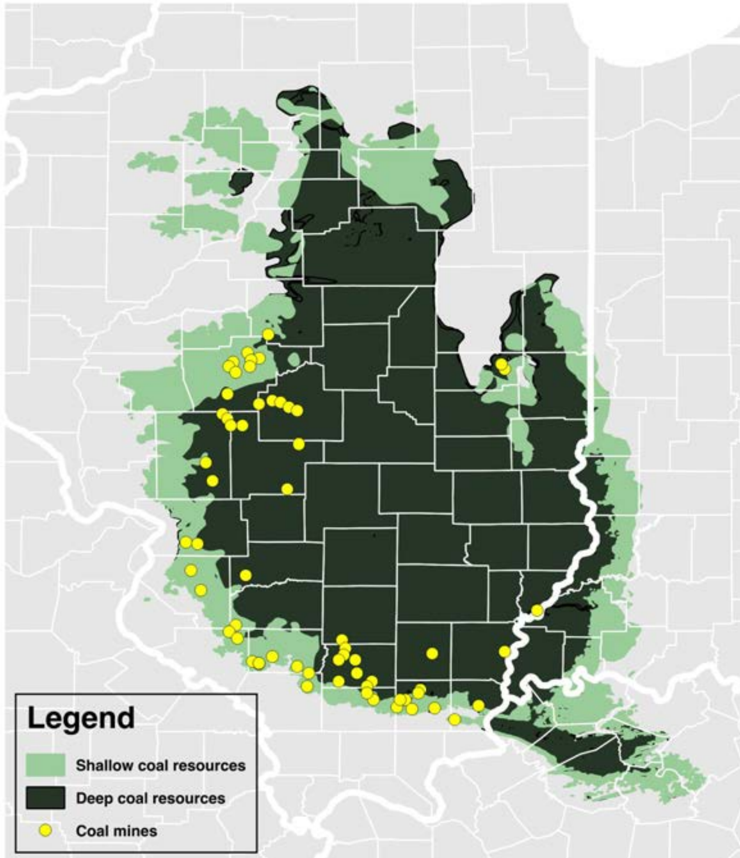
Within the Illinois Basin, I focus on coal mines whose spatial area exceeds the 95% percentile for that basin. Large coal mines have lower Hotelling, or scarcity rents, which is a potential endogenous component of the delivered coal price (Hotelling, 1931). Figure 3 shows the location of these mines over the Illinois Basin. On the demand side, I restrict my sample of counties to those whose distance from centroid to nearest Illinois coal resource is (i) less than the distance to the nearest Appalachian coal resource and (ii) less than 250 miles. The first restriction reduces the influence of coal from nearby Appalachian Basin on delivered coal prices. The second restriction reduces the influence from other coal basins. The resulting sample of counties is shown in Figure A.3.

[11] In the 1990s, the Illinois Coal Basin provided 20% of bituminous coal in the U.S. (Energy Information Administration, 1994) That share is likely to have been higher earlier in the 20th century before the large scale extraction of western U.S. coal.

[12] Historical coal mine data for other basins are still held by state-level natural resource agencies. These agencies typically possess a coal mine "final map" which dates a mine's closing, but not opening year.

[13] While the NCRA provides shape files of modern coal resources, much of the Illinois Basin has been characterized since the start of the 20th century such that one can interpret Figure 3 as indicating the location of coal resources known across much of the 20th century. For example, the shape of the Illinois Basin shown in Figure 3 matches closely with that found in Campbell (1908).

Figure 3: Location of Illinois Basin coal resources and mines



NOTES: Green (lighter) shaded area indicates shallow coal resources (< 200 ft. underground). Black (darker) shaded area indicates deep coal resources (> 200 ft. underground). Yellow dots show all large coal mines that operated at any point after 1890. County and state boundaries also shown.

Using the NCRA spatial data, I identify a county's distance to its nearest coal mine for each decade from 1890 to 1990 (see Appendix A for data construction details). Figure A.4 maps county distance to nearest mine from 1890-1950 showing how its spatial distribution has changed over the 20th century. From this data, I can also determine the decade in which a county's nearest mine switches from a shallow to a deep coal mine for the first time. Figure A.5 shows the timing of this switch for each sample county, stacked according to the decade when it occurs. Observe that the timing of the switching event differs across counties. 41% of sample counties experiences the switch during the 1900s; 37% experiences the switch during the 1960s.

I turn to three pieces of auxiliary evidence to verify the construction of my distance-based measure. First, the use of transport distance assumes that transport costs are an important component of delivered coal prices. Figure A.6 shows that transport costs were between 40-60% of national U.S. delivered coal prices over the first half of the 20th century. This is consistent with previous research emphasizing the high costs of transporting coal, the heaviest fossil fuel by heat content (Joskow, 1987; Busse and Keohane, 2007; McNerney, Farmer and Trancik, 2011; Preonas, 2018).

Second, my construction procedure relies on the Herfindahl Principle, in which power plants buy coal from its nearest mine. Using coal procurement data from 1990-1999 FERC-423 forms, Figure A.7 shows how the share of coal purchased by a county varies according to the rank in the transport distance to coal supplying counties.[14] Nearly 50% of purchased coal came from the nearest coal supplying county. There is a steep and sustained drop in the share of coal purchased from other counties.

As a final check, I regress observed county-level delivered coal prices obtained from the FERC-423 forms during the 1970s, 1980s, and 1990s against my distance-based measure for sample counties, conditional on state fixed effects. A perfect statistical fit is not expected nor desired as observed prices may contain endogenous components in addition to transport costs. The correlations shown in Table A.2 are statistically significant at the 5% level across all three decades and as such provide confidence that my distance measure captures variation in delivered coal prices.

4.2 Electricity capital

Fuel-specific electricity capital, or capacity, at or below the county level throughout the 20th century is also not directly available (see Appendix B). Instead, I turn to modern EIA-860 forms to construct a county-by-decade panel of relative coal capital from 1890 to 1990. Importantly, this construction of historical data is made possible because EIA-860 collects data on capacity, operating years, and primary fuel input of active and retired generating units at power plants that were operating at the time of reporting. The availability of retired generating units, in particular, enables one to observe historical electricity capital that is no longer active today (see Appendix A for construction details).

Three assumptions must be satisfied for my constructed relative coal capital to match historical values. First, all power plants since 1890 must continue to have at least one active generating unit today. If an entire power plant retires, their generating units would not appear in modern EIA-860 forms. To test the degree in which my constructed data may be missing retired power plants, I turn to available national historical data since 1920 (U.S. Census Bureau, 1975). Panel A of Figure A.8 compares U.S. electricity capacity (in gigawatt, GW) summed across generating units burning fossil fuels (i.e., coal, oil, and natural gas) constructed from EIA-860 forms against observed values from the U.S. Historical Census for

[14] Unfortunately, FERC-423 forms provide the county of origin for delivered coal, and not mine. Because one is unable to directly link power plants with coal mines, Figure A.7 provides a noisy test of the Herfindahl Principle.

the 1920-1970 period. Panel B of Figure A.8 provides a similar comparison but for annual capacity changes. While my constructed data under reports fossil fuel capacity levels prior to 1955, the two national data series track closely in terms of annual changes throughout the 1920-1970 period. This suggests that my constructed data may be missing power plants that began operation prior to 1920 but less likely those operating after 1920. Panels C and D of Figure A.8 draws a similar conclusion for electricity capital using hydro power.

My second and third assumptions are that a generating unit must not change its capacity and primary fuel during its lifetime. Section 6.1 discusses engineering reasons for why these features are likely to be stable over time. Nonetheless, such changes may occur. Table A.3 examines the consistency of key generating unit characteristics across the 1990-2012 EIA-860 forms. For each characteristic across columns of Table A.3, row values indicate the percentage of 1990-2011 EIA-860 forms with values that differed from that reported in the 2012 EIA-860 form. 75%, 94%, 97% and 80% of generating units consistently reported using the same capacity, primary fuel, opening year, and retirement year in 1990-2011 as was reported in 2012.

To examine the consistency of generating unit characteristics over a longer time horizon, I digitized the 1980 EIA "Inventory of Power Plants in the United States," the earliest available comprehensive generating-unit dataset, with data collected during the late 1970s. Figure A.9 plots generating unit 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.4 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 generating units with 92% of generating units that reported using coal in the late 1970s also reporting coal use in 2012.

Imputing missing small power plants The EIA-860 forms have one additional limitation: they exclude generating units on power plants with less than 1 megawatt (MW) of combined electricity capacity. This omission can be consequential for relative coal capital and how it is modeled. Suppose a county's non-coal electricity only comes from such small power plants. Then my data would erroneously assigned a zero value to non-coal capital, leading to a missing value for relative coal capital. Such data censuring would result in a smaller sample and a lower sample mean for relative coal capital. Similarly, if a county's coal-fired electricity is produced only by missing small power plants, then both coal capital and relative coal capital would be incorrectly assigned zero values. This form of censuring increases the skewness of relative coal capital.

How many power plants fall below the 1 MW threshold? In the absence of historical data, one can try to predict the frequency of such plants using the observed power plant size distribution. Figure A.10 plots the size distribution for all U.S. power plants appearing in the EIA-860 forms built during the 1910s and 1950s.[15] The solid and dashed lines show the fitted and predicted frequency of power plants between 1 and 30 MW capacity using a flexible polynomial specification. Because of the skewness of the size distribution, Figure A.10 predicts there are more power plants below 1 MW built each decade across the U.S. than power plants of any other size. Figure A.10 also suggests suggests a "downscaling" method for imputing missing small power plants. As detailed in Appendix C, my imputation procedure takes the predicted national frequency of new missing power plants for each decade, splits it by fuel according to the national coal share, and assigns new power plant capital uniformly across electricity-producing counties.

Column 1 of Table A.5 shows summary statistics for raw relative coal capital. Of the 2,371 observations in my county-by-decade sample, 1,248 observations are missing relative coal capital. 825 observations have zero relative coal capital values. Columns 2-4 provide summary statistics for relative coal capital following the imputation procedure mentioned above where the national power plant size distribution for each decade is fitted with 3rd, 4th, and 5th order polynomial functions, respectively. As expected, my imputation procedure increases the sample size, raises the mean, and reduces the skewness of relative coal capital. Column 5 displays summary statistics for relative coal capital using an alternative imputation approach which simply adds 1 MW to new coal and non-coal capital investment in each county-by-decade observation. Robustness checks in Section 5.2 will examine all five measures of relative coal capital in Table A.5.

Finally, regardless of my imputation procedure, relative coal capital, remains right-skewed. This skewness in my outcome variable can be mitigated by the log transformation applied in equation (4). However, a log transformation is also sensitive to small values of the untransformed variable, which may result from my imputation procedure. To produce estimates that are less sensitive to my imputation method, I estimate equation (4) using a Poisson model.[16] I also consider other models as robustness checks in Section 5.2.

[15]Size distributions for plants built during other decades have similar shapes.

[16]Specifically, the Poisson version of equation (4) is

$$\widetilde{K}_{it} = exp\left(\sum_{\substack{-2\leq\tau\leq10\\ \tau\neq0}} \beta^{\tau}[\ln d_i^0 \times \mathbf{1}(\tau = h)] + \pi \ln d_{it} + \sum_{\substack{-2\leq\tau\leq10\\ \tau\neq0}} \gamma^{\tau}\mathbf{1}(\tau = h) + \rho_i + \phi_{st}\right) + \mu_{ist} \tag{4'}$$

The Poisson model has the additional benefit of 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.

5 Reduced-form results

This section presents reduced-form evidence of path dependence in the coal composition of electricity capital. I first discuss baseline estimates before turning to robustness checks.

5.1 Baseline estimates of path dependence

The thick solid red line in Figure 4 shows my baseline point estimates of β^τ from equation (4) estimated using a Poisson model with relative coal capital, $\tilde{K}_{it} = \frac{K_{cit}}{K_{nit}}$, as the outcome. These estimates are also printed in column 1 of Table 1. Following the discussion in Section 3.3, I include event time, county, and state-by-decade fixed effects to control for unobserved local coal demand and supply shocks. To further ensure a path dependence interpretation in the case that local coal prices are serial correlated, I also control for a county's distance to the current nearest mine as a proxy for contemporaneous coal prices.

Standard Poisson models impose that the first and second moments of the outcome be equal. Table A.5 shows that the variance of relative coal capital exceeds its mean. To address this overdispersion issue, my baseline model has standard errors clustered at the county level. This adjustment relaxes the assumption of equal first and second moments by allowing arbitrary forms of within-county heteroskedasticity and serial correlation in the error term. The darker shaded area of Figure 4 shows the 95% point confidence intervals for β^τ using county-level clustered standard errors. As an alternative, the lighter shaded area shows the 95% point confidence interval when errors are clustered at the state-by-decade level, which allows arbitrary heteroskedasticity and spatial correlation across counties in the same state and decade. These two confidence intervals cover similar ranges.

Several features of my baseline estimates are worth highlighting. First, I do not detect lead effects (i.e., $\beta^\tau : \tau < 0$). The absence of differential pre-event trends in relative coal capital supports my contemporaneous exogeneity assumption. Second, I detect statistically significant lagged effects, all except for the 1 decade lag effect at the 1% level (i.e, $\beta^\tau : \tau > 0$). Two features of these lagged effects suggest that past coal price shocks are altering the subsequent fuel imbalance of new electricity capital investments. First, these lagged effects become more negative over time. Dynamics driven only by the depreciation of previously built capital should dissipate, not strengthen, following a coal price shock. Second, observe that the shape of the lagged effects shown in Figure 4 exhibits two distinct jumps after the switching event: a 1% increase in coal prices (as approximated by distance to the shallow mine) discretely lowers relative coal capital by 4.11% and 6.20% two and seven decades later, respectively. Given that the average lifespan of generating units in my sample is 4.7 decades, the timing of these jumps roughly coincides with the expected construction of two

Table 1: Baseline reduced-form estimates of path dependence

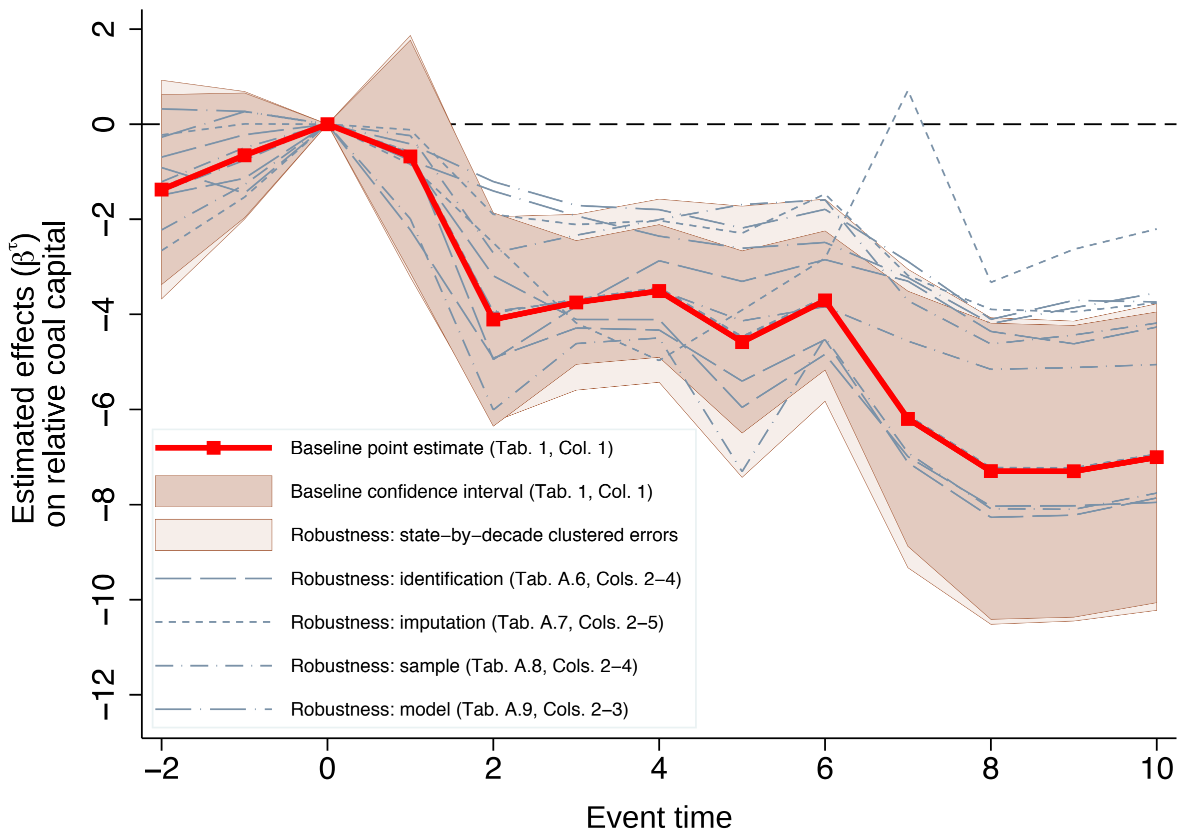
	(1) relative coal capital	(2) relative coal capital investment
$\ln d_i^0$ (β^τ)		
2 decades lead	-1.38	0.60
	(1.02)	(2.00)
1 decade lead	-0.66	-0.33
	(0.67)	(2.32)
	–	–
1 decade lag	-0.68	-4.47**
	(1.25)	(1.89)
2 decades lag	-4.11***	-3.46
	(1.15)	(2.17)
3 decades lag	-3.75***	-7.75**
	(0.66)	(3.61)
4 decades lag	-3.51***	-5.68
	(0.71)	(5.21)
5 decades lag	-4.58***	-4.46
	(0.98)	(3.59)
6 decades lag	-3.71***	-2.93
	(0.75)	(3.89)
7 decades lag	-6.20***	-10.1**
	(1.37)	(4.20)
8 decades lag	-7.30***	-6.42
	(1.59)	(4.27)
9 decades lag	-7.30***	-3.83
	(1.56)	(4.55)
10 decades lag	-7.01***	-2.77
	(1.56)	(5.82)
$\ln d_{it}$ (π)	-1.53***	-3.48***
	(0.53)	(0.88)
Observations	2,371	2,371
Counties	261	261

NOTES: Estimates of β^τ and π from equation (4) using Poisson model. Outcomes variables are at the county-by-decade level. Each model includes event time, county, and state-by-decade fixed effects. County sample shown in Figure A.3. Time period is 1890-1990. Outcome in column 1 is relative coal capital. Outcome in column 2 is relative coal capital investment. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

subsequent vintages of electricity capital. For a sense of relative magnitude, one can compare lagged and contemporaneous effects (i.e., π) of coal price shocks. Lagged effects two and seven decades later are 2.7 (i.e., $-4.11/-1.53$) and 4.1 (i.e., $-6.20/-1.53$) times that of the contemporaneous effect, respectively.

As further confirmation that past coal price shocks are affecting the fuel-composition of capital investments, column 2 of Table 1 shows estimates of β^{τ} when the outcome is relative coal capital investment, $\tilde{X}_{it} = \frac{X_{cit}}{X_{nit}}$, the ratio of new coal to non-coal electricity capital built in each county i and decade t. Coal price shocks alter relative coal capital investment over many subsequent decades.

Figure 4: Reduced-form estimates of path dependence in relative coal capital



Notes: Thick solid red line and darker shaded area show baseline point estimates and 95% confidence intervals for β^{τ} from equation (4) with county-level clustered standard errors. Outcome is relative coal capital at the county-by-decade level. County sample shown in Figure A.3. Time period is 1890-1990. Estimates also shown in column 2 of Table 1. Lighter shaded area shows the 95% confidence interval when estimating equation (4) with state-by-decade clustered standard errors. Thinner non-solid blue lines show point estimates from various robustness tests displayed in columns 2-4 of Table A.6, columns 2-5 of Table A.7, columns 2-4 of Table A.8, and columns 2-3 of Table A.9.

5.2 Robustness checks

I turn now to a series of robustness checks. These tests are designed to examine the robustness of my baseline results to remaining identification concerns as well as to data imputation, sample restriction, and statistical modeling considerations. Point estimates for all robustness checks are shown as thin non-solid blue lines in Figure 4.

Remaining identification concerns Table A.6 examines remaining potential identification concerns. Column 1 replicates my baseline estimates. To explicitly control for electricity demand from the manufacturing and residential sectors, column 2 augments my baseline model with three county-by-decade covariates that are available throughout the 20th century: population, number of manufacturing establishments, and manufacturing employment, all in logs (see Appendix A for data details).

On the supply-side, distance to the nearest coal mine is only one component of delivered coal prices. Richer models of spatial competition allow for other components that may be correlated with distance to the nearest mine. One such component is the mill price set by the nearest mine which can reflect its marginal cost of coal extraction (Hotelling, 1929; C. d'Aspremont, 1979; Salop, 1979; Vogel, 2008) and the Hotelling rent associated with the size of its coal resource (Hotelling, 1931; Gaudet, Moreaux and Salant, 2001). The mill price may also reflect heterogeneity in coal quality such as in heat, ash, and sulfur content, which dictates the amount of coal needed to produce a unit of electricity. To control for mine-specific heterogeneity, column 3 augments my baseline model with a fixed effect for the nearest shallow mine right before the initial switch to deep coal.[17]

Finally, if coal mines can price discriminate, the delivered coal price may also include county-by-mine specific mark-ups. For example, if mine A is the nearest supplier to a county in one period but becomes the second nearest supplier in the next period following the opening of mine B, mine B could price-discriminate by setting a buyer's coal price equal to its distance to mine A (MacLeod, Norman and Thisse, 1988; Vogel, 2011). Under this form of market power, distance to mine A would directly determine a county's subsequent coal price even if mine A were no longer supplying coal.[18] To remove the possibility of price-discriminating from affecting my results, column 4 estimates my baseline model after removing the 42 counties for which earlier supplying shallow mine becomes the second nearest mine in any decade after the switching event.

[17] There are 52 mines that served as the nearest supplier to sample counties right before the switching event.

[18] There is evidence from recent years showing price discriminating behavior by railroad companies that sell coal to power plants (Busse and Keohane, 2007; Preonas, 2018). Little is known, however, about whether such mark-ups existed across the longer time period considered in this study.

Across columns 2-4 of Table A.7, lead coefficients and lagged coefficients for up to five decades are similar to my baseline estimates both in terms of point estimates and standard errors. Lagged coefficients beyond five decades are also similar in magnitude in columns 2 and 4. For column 3, longer lagged effects are smaller in magnitude though their point estimates still fall within the 95% confidence interval of my baseline estimates and their standard errors are statistically significant.

Imputing missing small power plants Table A.7 examines the sensitivity of my estimates to how missing small power plants are imputed. Column 1 replicates my baseline result using a 4th order flexible polynomial function to fit the national size distribution of new power plants built each decade. Columns 2 and 3 employ the same imputation procedure but uses 3rd and 5th order polynomial functions, respectively. My point estimates and standard errors are little affected. In column 4, I use a simpler, though less informed, imputation procedure by adding 1 MW capacity to coal and non-coal capital investment. The overall shape and precision of lagged effects are unchanged. However, the magnitude of these coefficients are smaller, which likely reflects the smaller contemporaneous coal price effect, π. Finally, column 5 estimates equation (4) using unadjusted relative coal capital. As expected, the estimating sample is much smaller. Despite that, lagged effects up to 5 decades exhibit magnitudes that are similar to my baseline estimates. 3 to 5 decade lagged effects are also statistically significant at the 10% level. Longer lagged effects, however, are noisier, likely because of insufficient statistical power.

Estimating sample Table A.8 considers different county sample definitions. For my baseline result, reproduced in column 1, the sample is restricted to counties within 250 miles of the nearest Illinois Basin coal resource 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 similar point estimates and standard errors. 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 smaller possibly due to the influence of other coal resources. However, these point estimates are within the confidence intervals of my baseline estimates.

Modeling choices The log-log functional form of equation (4) implicitly assumes that the relationship between relative coal capital and coal prices is an isoelastic function. While this is suitable for the theory to be considered in Section 6, this functional form assumption may not be empirically supported. Figure A.11 examines whether the data supports linearity by estimating a variant of equation (4) that breaks log distance to the nearest mine into discrete bins, and allows a flexible relationship between distance and relative capital for each event-time period. Figure A.11 shows the log relative coal capital predicted by log distance to the contemporaneous nearest mine, and by the log distance to the earlier shallow mine two and seven decades after the switching event. Linearity appears to be a reasonable assumption for all three effects.

I also consider two alternatives to the Poisson model. The first alternative is a log-log linear model. The downside to such an approach is that a log transformation is sensitive to small values of an outcome variable, which may arise from how missing small power plants are imputed. A second alternative is to employ a negative binomial model. In contrast to a Poisson model which semi-parametrically addresses overdispersion via clustered standard errors, a negative binomial model parametrically fits the variance of the outcome as a function of the expected mean outcome.

Table A.9 considers both alternative models, with column 1 reproducing my baseline Poisson model estimates. Column 2 estimates equation (4) using a log-log linear model, while column 3 uses a negative binomial model. Both alternative models produce statistically significant lagged effects, mostly at the 5% level. The overall shape of lagged effects from these two models, including the jumps at two and seven decades later, also mirrors that of column 1. However, the magnitude of these lagged effects are smaller when using these alternative models, which is reflected also by smaller contemporaneous coal price effects.

6 Mechanisms

Section 5 finds reduced-form evidence that temporary coal price shocks have increasingly negative effects on relative coal capital that lasts decades after the shock. Different mechanisms can generate such dynamics. Further evidence on the relevant mechanism would allow for a structural interpretation of my reduced-form results.

Prior literature highlights two possible mechanisms in the electricity sector: increasing returns to scale and the accumulation of fuel-specific productivity. Before examining these two channels, this section first briefly summarizes three pertinent features of the electricity sector. I then introduce a multi-sector model of the electricity sector capturing these features. This model informs a series of empirical tests designed to isolate the relevant mechanism.

Finally, I provide additional evidence testing for the presence of other possible mechanisms.

6.1 Features of an electric power plant

Fixed input proportions To produce electricity, most power plants must first produce steam by burning fuel. This is done in a boiler. Boilers are typically designed to consume a particular primary fuel at a particular rate (Avallone, Baumeister and Sadegh (2006), p. 875). As a consequence, sustained use of other fuels or use of the primary fuel at other quantities can often result in large efficiency losses (Avallone, Baumeister and Sadegh (2006), p.871). This implies that boilers, once built, can only employ inputs in fixed proportions. This "clay"-like nature of boilers has long been recognized in the energy economics literature (Komiya, 1962; Joskow, 1985, 1987; Atkeson and Kehoe, 1999; Fabrizio, Rose and Wolfram, 2007).

Returns to scale Boilers supply steam to generators for conversion into electricity. When a new boiler is installed, it can serve both new and existing generators, providing efficiency gains for generating units across multiple vintages. This spillover effect, together with newer and larger boilers typically being more thermally efficient, forms a physical basis for increasing returns to scale. Nerlove (1963) and Christensen and Greene (1976) provide seminal early estimates of increasing returns to scale in the electricity sector.

Imperfect substitutability across fuel-specific electricity Electricity produced from different fuels often exhibit different characteristics, making them imperfect substitutes. For example, coal-fired generating units provide "base load" electricity that can meet a steady level of demand. However, such units cannot be easily ramped up or down to meet variable demand as readily as natural gas-fired generating units. Similarly, when considering future energy transitions, electricity generated from solar and wind may be less reliable than electricity generated from fossil fuels due to the intermittent availability of these renewable resources. This imperfect substitutability across fuels is a crucial element of the directed technical change model by Acemoglu et al. (2012).

6.2 A model of structural change in the electricity sector

I consider a model of the electricity sector that nests these three features: (i) vintages of electricity capital with fixed factor productivities, (ii) increasing returns to scale across capital vintages, and (iii) imperfect substitutability between electricity from coal and from other fuels. Each feature is represented by a separate tier in the production function.

The index t denotes the time increment between each capital vintage. At the top tier, the final good, electricity, Y_t, is produced using two intermediate goods, Y_{ct} and Y_{nt}, representing electricity from coal and all other fuels respectively.[19] Specifically, it takes the following Constant Elasticity of Substitution form

$$Y_t = \left(Y_{ct}^{(\epsilon-1)/\epsilon} + Y_{nt}^{(\epsilon-1)/\epsilon}\right)^{\epsilon/(\epsilon-1)} \tag{5}$$

where ϵ is the elasticity of substitution between electricity produced from the two intermediate sectors. I assume that electricity produced by different fuels are imperfect substitutes, $\epsilon \in (1, +\infty)$. The price of the final good is normalized to 1.

Fuel-specific electricity, the intermediate good, is produced via middle and lower tiers of the production structure. These two tiers combine to form the following expression

$$Y_{jt} = (\min[A_{Xjt}X_{jt}, A_{Ejt}E_{jt}])^{\alpha}(\min[A_{Xjt-1}(1-\delta)X_{jt-1}, A_{Ejt-1}E_{jt-1}])^{\alpha} \quad \text{for} \quad \text{j}\in\{\text{c,n}\} \tag{6}$$

where δ is the capital depreciation rate and $\alpha \in \{0, 1\}$ is the fuel-specific electricity elasticity of input. In the middle tier, fuel-specific electricity is generated by combining generating units from two vintages in a Cobb-Douglas function with scale parameter $\psi = 2\alpha$.[20] At the lowest tier, generating units combine t-vintaged capital, X_{jt}, and fuel, E_{jt}, in fixed proportions in a Leontief function that captures the clay-like nature of boilers. These fixed proportions are A_{Xjt} and A_{Ejt}, which represent capital and fuel productivities, respectively.

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 tier imply $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ and $A_{Xjt-1}(1-\delta)X_{jt-1} = A_{Ejt-1}E_{jt-1}$ for each vintage t and fuel j. Next, the power plant's first order conditions in period t, rewritten in terms of current-vintage relative coal capital investment, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$, is (see Appendix D.1 for full derivation)

$$\widetilde{X}_t = \widetilde{w}_t^{\frac{\epsilon}{\varphi-1}} \widetilde{X}_{t-1}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \widetilde{A}_{Xt}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \tag{7}$$

where $\widetilde{w}_t = \frac{w_{ct}}{w_{nt}}$ is the relative input price index (see Appendix D.1) and $\widetilde{A}_{Xt} = \frac{A_{Xct}}{A_{Xnt}}$ is the ratio of capital productivity for coal and non-coal generating units of vintage t. $\varphi = (1-\alpha)(1-\epsilon) < 0$, from earlier assumptions. Equation (7) provides two channels through

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

[20] This allows for diminishing marginal product under varying returns to scale. Otherwise, the relative input demand curve becomes upward sloping. The assumption that returns to scale is constant for coal and non-coal electricity production is examined in Table 2.

which past relative input prices, $\widetilde{w}_{t-1}$, affect current-vintage relative coal capital investment, $\widetilde{X}_t$. First, it is evident from applying equation (7) recursively that past relative input prices affect past-vintage relative coal capital investment, $\widetilde{X}_{t-1}$. This is the scale effect. Second, while not explicitly modeled here, in the presence of sector-biased technical change, past relative input prices could also affect current-vintage relative capital productivities, $\widetilde{A}_{Xt}$. This occurs if there is accumulation over time in fuel-specific capital productivity such as via learning-by-doing. This channel is the productivity effect.

To empirically isolate which of these two channels drives my path dependence results in Section 5, I turn next to a series of nested empirical tests informed by the tiered structure of my model. First, I conduct power plant-level cost regressions for plants that only use coal to recover the plant-level scale parameter. Because such estimates may be biased in the presence of productivity effects, I then turn to generating unit-level regressions to test for productivity effects.

6.3 Testing for scale effects at the power plant level

To recover the scale parameter, ψ, I follow the approach developed initially by Nerlove (1963) and implemented by Christensen and Greene (1976) to estimate returns to scale in the electricity sector. I use power plant-level cost data from the Utility Data Institute (UDI) from 1981-1999. To remove the influence of the elasticity of substitution parameter, ϵ, I restrict my sample to power plants p in county i and state s that exclusively use coal. Cost minimization of equation 6 implies the following regression of non-fuel cost (see Appendix D.2 for full derivation)

$$\ln \overline{\text{non\_fuel\_cost}}_{pis} = \frac{1}{\psi} \ln \overline{Y}_{pis} + \theta' \overline{\boldsymbol{Z}}_{pis} + \eta_{pis} \tag{8}$$

where the bar indicates time-averaged variables over 1981-1999. My parameter of interest is the scale parameter ψ. $\overline{Y}_{pis}$ is electricity output in megawatt-hours (MWh). $\overline{\boldsymbol{Z}}_{pis}$ is a vector of cross-sectional controls. They include observed log power plant-level delivered coal price from UDI, state fixed effects and the latitude and longitude of the county centroid. I also control for differences across transmission grids by including NERC region fixed effects. Standard errors, η_{pis}, are clustered at the county level.

Table 2 displays estimates of ψ for coal-only power plants in my baseline county sample. I estimate a scale parameter of $\psi = 1.8$ that is statistically significant at the 1% level. Potential simultaneity bias in equation (8) has been noted as early as Nerlove (1963). In particular, electricity prices for electric utilities are historically regulated to cover a plant's average costs. As a consequence, electricity output may be correlated with unobserved determinants of non-fuel costs. To address this endogeneity concern, I use past delivered coal prices as an instrument for current electricity output via an instrumental variables (IV) approach. Specifically, my instrument is the interaction between county distance to the nearest shallow mine before the switch to deep coal and the number of decades since that switching event. For identification to be valid, my distance-based proxy for past coal prices must affect current non-fuel costs only through current output. Specifically, my first stage regression is

$$\ln \overline{Y}_{pis} = \kappa_1 \ln d_i^o * sinceEvent_i + \kappa_2 \ln d_i^o + \kappa_3 sinceEvent_i + \vartheta' \overline{\boldsymbol{Z}}_{pis} + \nu_{pis} \tag{9}$$

Equation (9) estimates the event time-varying effects of past coal price shocks and is the cross-sectional analog to my panel estimator in equation (4). In particular, κ_1 captures the event time-varying effect of past coal price shocks, or the slope of the lagged effects shown in Figure 4. Column 2 of Table 2 shows an IV estimate that is statistically significant at the 1% level and similar in magnitude to my OLS estimate. Furthermore, this IV estimate is robust to the potential presence of a weak instrument. The p-value and confidence interval from a conditional likelihood ratio test strongly reject a null hypothesis that the coefficient on electricity output in equation (8) is zero (Moreira, 2003). Henceforth, my preferred scale parameter estimate is $\psi = 1.66$ from column 2.

The production function presented in Section 6.2 assumes that coal and non-coal electricity exhibit the same returns to scale. To examine this assumption, I estimate equation (8) for power plants that consume natural gas and oil in columns 3 and 4 of Table 2. Because the UDI database has few plants that only consume natural gas and oil in my county sample, for adequate statistical power, I expand the sample to cover the entire U.S. and to plants UDI designates as consuming natural gas and oil as primary fuels.[21] I estimate scale parameters for gas and oil-fired power plants that are statistically indistinguishable from that of coal-fired power plants.

Finally, despite differences in sample time periods, it is noteworthy that the scale parameter estimates from 1981-1999 in Table 2 are similar to estimates from the prior literature. Nerlove (1963)'s study of electricity sector firms in 1955 recovers a scale parameter of 1.39. For the same sample of firms, Christensen and Greene (1976) estimate an analogous scale parameter of 1.26.

[21] Federal forms used in compiling the UDI data do not explicitly designate a power plant's primary or secondary fuel. For multi-fuel plants, UDI establishes primary fuel by calculating the energy input of each fuel consumed and then assigning primary fuel to the fuel with the highest energy input.

Table 2: Returns to scale regressions at the power plant-level

	(1)	(2)	(3)	(4)
	Outcome is ln non-fuel cost			
$\ln \overline{Y}_{pis}$ $(1/\psi)$	0.56***	0.60***	0.57***	0.51***
	(0.028)	(0.087)	(0.052)	(0.054)
CLR p-value		0.0081		
CLR confidence int (90%)		[.54, .89]		
Implied scale parameter, ψ	1.78***	1.66***	1.75***	1.96***
	(0.090)	(0.24)	(0.16)	(0.21)
Model	OLS	IV	OLS	OLS
County sample	Baseline	Baseline	All U.S.	All U.S.
Fuel input	Only coal	Only coal	Primary gas	Primary oil
Power plants	103	96	32	73

NOTES: Estimates of ψ from equation (8) using power plant-level log non-fuel cost as outcome. All models include observed power plant-level log fuel price, state and NERC region fixed effects, and county centroid longitude and latitude. Non-fuel cost and fuel price are 1981-1999 averages. Columns 1 and 2 include coal-only power plants in baseline county sample shown in Figure A.3. Columns 3 and 4 include all U.S. power plants using natural gas and oil as primary fuels, respectively. Columns 1, 3, and 4 show results from OLS regressions. Column 2 shows the result from an IV regression with equation (9) as the first stage. Column 2 also shows the p-value and confidence interval (in brackets) from a conditional likelihood ratio test against a null hypothesis that 1ψ is zero. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

6.4 Testing for productivity effects at the generating unit level

In the presence of endogenous technical change, past coal prices would also directly affect the accumulative of coal-specific productivity. If so, the IV estimator in Section 6.3 would not satisfy the exclusion restriction needed for identifying the scale parameter. To detect productivity effects, I turn to generating unit-level regressions where productivity effects can be most cleanly isolated. I employ two standard measures of generating-unit productivity. Following Davis and Wolfram (2012), my measure of capital productivity, A_{Xct}, is the ratio of electricity output to generating unit capacity on an annualized basis. My second measure is thermal efficiency, or the ratio of heat from electricity produced to heat from fuel consumed. This measure corresponds to fuel productivity, A_{Ect}. To obtain both measures, I combine the previously mentioned generating unit capacity data from EIA-860 forms with generating-unit electricity production and boiler-level fuel consumption from EIA-923 forms (see Appendix A for data details). Averaging generating unit-level data across 2009-2012,[22] I estimate the

[22] Because productivity accumulates over time, any lagged effects from past coal price shocks would be stronger and more detectable using recent productivity data.

following regression for generating unit g, in power plant p, located in county i and state s using both productivity measures as outcome

$$\ln \overline{A}_{gpis} = \omega_1 \ln d_i^o \times sinceEvent_i + \omega_2 \ln d_i^o + \omega_3 sinceEvent_i + \lambda' \overline{\boldsymbol{Z}}_{gpis} + \upsilon_{gpis} \tag{10}$$

where the set of controls $\overline{\boldsymbol{Z}}_{gpis}$ includes state and NERC region fixed effects, and the latitude and longitude of the county centroid. The standard error, υ_{gpis}, is clustered at the county level. As with equation 9, ω_1 in equation 10 captures the event time-varying effects of past coal price shocks on generating unit productivity.

Table 3: Productivity regressions at the generating unit-level

	(1)	(2)	(3)	(4)
	Outcome is ln			
	capital productivity		fuel productivity	
$\ln d_i^o \times sinceEvent_i$ (ω_1)	-0.052	-0.032	0.015	0.034
	(0.032)	(0.031)	(0.047)	(0.047)
$\ln d_i^o$ (ω_2)	0.12	0.078	-0.33	-0.37
	(0.17)	(0.18)	(0.26)	(0.27)
$sinceEvent_i$ (ω_3)	0.14	0.034	-0.12	-0.22
	(0.12)	(0.11)	(0.18)	(0.16)
age		-0.012**		-0.011
		(0.0048)		(0.0072)
Generating units	224	224	224	224

NOTES: Estimates from equation 10 using generating unit-level outcomes. All models includes state and NERC region fixed effects, and county centroid longitude and latitude. Sample includes generating units in coal-only power plants located in baseline county sample shown in Figure A.3. Outcome in columns 1 and 2 is the 2009-2012 average ratio of electricity generation to generating unit capital, which approximates capital productivity. Outcome in columns 3 and 4 is the 2009-2012 average thermal efficiency, which approximates fuel productivity. Models in columns 2 and 4 additional include generating unit age. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Results are shown in Table 3 for generating units in coal-only power plants located in my baseline county sample. Columns 1 and 2 examine log capital productivity. I do not detect statistically significant effects of past coal prices. Imprecise effects become smaller when I additionally control for for the age of the generating unit in column 2. Columns 3 and 4 replicate columns 1 and 2 but with log thermal efficiency as the outcome. Again, I do not find that past coal price shocks affect subsequent fuel productivity.

It is perhaps unsurprising that local productivity does not respond to coal price shocks at the county level. Most known determinants of productivity operate at broader scales.

For example, innovation activity responds to incentives in the research sector, which are rarely confined to individual counties. Similarly, if productivity improvements occur through human capital accumulation, such effects are unlikely to be restricted to a single county when firms with several power plants can distribute human capital across multiple counties.

6.5 Alternative mechanisms

I now test for other mechanisms that fall outside my model in Section 6.2.

Electricity sector regulation Capital investment responds to regulation. In particular, two regulations may be pertinent for electricity capital decisions. For much of the 20th century, electricity producers faced cost-of-service regulation by state Public Utility Commissions (PUC) which set output prices to ensure producers recover "prudently incurred" variable costs plus a regulated rate of return on capital investments. As noted first by Averch and Johnson (1962), when the allowed rate of return exceeds the cost of capital, electricity producers have an incentive to inefficiently over invest in capital relative to other inputs.[23] Conceptually, the Averch-Johnson effect is unlikely to drive my reduced-form results in Section 5. Structural change pertains to the composition of capital across different fuels and not the overall level of capital. Because cost-of-service regulations for electricity producers do not specify different rates of return for electricity capital using different fuels, such regulations are unlikely to alter the relative marginal product of capital across fuels. This is supported empirically by column 2 of Table A.10 which estimates my baseline model in equation (4) but using only the subset of observations during which there is no state PUC regulation of electric utilities (see Appendix A for details). Despite the smaller sample size, I still detect statistically significant lagged effects with magnitudes that are similar to my baseline estimates in column 1 of Table A.10.

A second important regulation is the U.S. Clean Air Act. Beginning with the 1970 Clean Air Act Amendments, U.S. counties with concentrations of criteria air pollution exceeding national ambient air quality standards were labeled as being in "nonattainment". Both existing and new polluting facilities in nonattainment counties are required to invest in pollution abatement equipment. Because coal is dirtier than other fuels, it is possible that the 1970 and later Clean Air Act Amendments altered the fuel composition of electricity capital. Table A.10 provides two empirical tests. In column 3, I restrict the sample to observations during 1890-1960, the period before the introduction of the 1970 Clean Air Act Amendments.

[23] The literature has found some empirical support for the Averch-Johnson effect in the U.S. energy sector. Joskow and Rose (1989) reviews evidence from the 1960-1980 period. Cicala (2015) finds this effect using more recent data.

I find lagged effects that are similar in magnitude to my baseline results. Column 3 also indicates that my path dependence estimates are stable across the 20th century. To provide a more direct test, column 4 restricts my sample to the subset of counties that never received a nonattainment designation during the 20th century. Again, I do not find lagged effects that differ much from my baseline estimates. In summary, Table A.10 suggests that these two regulations are not driving my path dependence results.

Upstream and downstream sectors The electricity sector consumes fuel from upstream extraction and transport sectors and produces electricity for downstream manufacturing and residential sectors. Features of these up- and down-stream sectors could generate the path dependence detected in Section 5.

Power plants typically procure coal through long-term contracts with mines (Joskow, 1987). As a consequence, plants may continue buying coal from certain producers even as contemporaneous circumstances change. Joskow (1987) showed that the average coal contract length in 1979 lasted 12.8 years. More recent work by Kozhevnikova and Lange (2009) and Jha (2015) find this duration has since decreased to 4.4 years in the 1980s and 1990s. Coal contracts of such duration are unlikely to generate lagged effects over multiple decades, as detected in Section 5. Furthermore, even if coal contracts were of longer duration, it is unclear why they would cause lagged effects to strengthen over time.

Rail and highway networks serve as complementary capital for delivering coal to power plants. Increasing returns along with sunk costs in the transport sector can also generate path dependence in the fuel composition of the electricity sector.[24] To examine whether past coal price shocks affect the coal transport sector, I turn to a county-level version of the specification in equation (10). Column 1 of Table A.11 examines whether there are event time-varying effects of past coal price shocks on log railroad density in 2010 (in miles per square mile). I do not find such effects which is unsurprising given that most modern U.S. railroads lines were already established by the end of the 19th century (Atack, 2013). Column 2 of Table A.11 also fails to detect effects of past coal price shocks on log highway density in 2010 (in miles per square mile).

Finally, I consider downstream effects. Previous literature has detected long-term effects of historical access to hydropower electricity on local manufacturing sector employment (Kline and Moretti, 2014) and population density (Severnini, 2014). Indeed, one alternative explanation for my results involves downstream manufacturing sectors also exhibit increasing returns to scale and somehow preferring electricity produced from coal over that from

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

other fuels. If so, one would expect my lagged effects to be altered when I control for local manufacturing sector demand. Column 4 of Table A.6 shows that is not the case. Preference sorting by residential households provides another potential downstream mechanism (Tiebout, 1956). If households with low valuation for environmental amenities sort into historically coal-dependent locations, local residents may institute more lenient local environmental policies that continue the expansion of coal-fired electricity even as economic circumstances that initially favored coal disappear. Using the same specification as in columns 1 and 2, columns 3 and 4 examine this sorting mechanism by estimating event time-varying effects of past coal price shocks on the county share of the population that below to one of three major environmental NGOs in 1996 and the county share of eligible voters who voted for the Republican Presidential candidate in 2000 (see Appendix A for data details). Neither proxy for environmental preferences responds to past coal price shocks.

7 Structural interpretation

This section provides a structural interpretation for my reduced-form results. First, I formally define path dependence in energy transitions as a function of two parameters: returns to scale, ψ, and the elasticity of substitution between electricity produced from coal and other fuels, ϵ. Second, using this definition, I recover the elasticity of substitution implied by my reduced-form estimates of path dependence in Section 5 and by estimates of the scale parameter in Section 6.3. Finally, I explore, through simulations, the conditions under which a temporary policy intervention can induce a permanent energy transition away from coal.

7.1 Formal definitions of path dependence

In the absence of productivity accumulation effects and other possible mechanisms, I can now formally defined scale-driven path dependence within the model presented in Section 6.2. To do this, I first apply a log transformation to equation (7) and rewrite it recursively. Current-vintage relative coal capital investment becomes

$$\begin{aligned}\ln \widetilde{X}_t &= \frac{\epsilon}{(\varphi-1)}\ln \widetilde{w}_t + \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2}\ln \widetilde{w}_{t-1} + \frac{\alpha^2(1-\epsilon)^2\epsilon}{(\varphi-1)^3}\ln \widetilde{w}_{t-2} + \ldots \\ &\quad + \frac{\alpha(1-\epsilon)}{(\varphi-1)}\ln \widetilde{A}_{Xt} + \frac{\alpha^2(1-\epsilon)^2}{(\varphi-1)^2}\ln \widetilde{A}_{Xt-1} + \frac{\alpha^3(1-\epsilon)^3}{(\varphi-1)^3}\ln \widetilde{A}_{Xt-2} + \ldots \\ &= \sum_{s=0}^{\infty}\frac{\epsilon}{(\varphi-1)}\left[\frac{\alpha(1-\epsilon)}{(\varphi-1)}\right]^s \ln \widetilde{w}_{t-s} + \sum_{s=0}^{\infty}\left[\frac{\alpha(1-\epsilon)}{(\varphi-1)}\right]^{s+1}\ln \widetilde{A}_{Xt-s} \qquad (11)\end{aligned}$$

where s is a lagged time index. An increase in relative coal price lowers contemporaneous relative coal capital investment, $\frac{\partial \ln \widetilde{X}_t}{\partial \ln \widetilde{w}_t} < 0$, and also lowers future relative coal capital investment, $\frac{\partial \ln \widetilde{X}_t}{\partial \ln \widetilde{w}_{t-s}} < 0$. The relative magnitude of these two effects dictate the strength of path dependence in energy transitions. Formally,

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 the relative coal price triggers a downward shift in the relative marginal product of capital investment in subsequent periods. This shift is the net result of two countervailing forces. To illustrate these two forces, I consider each first in isolation.

The first force is increasing returns to scale, captured by the scale parameter ψ. Suppose there is only one intermediate sector which uses coal for electricity. When $\psi > 1$, the cross partial derivative of Y_{ct} with respect to past and current capital investment exceeds the second partial derivative of Y_{ct} with respect to current capital investment.[25] This "push" enables relative coal prices to have a stronger effect on future coal capital investment than on current coal capital investment.

However, when there is more than one sector, $\psi > 1$ alone does not dictate the strength of path dependence. A countervailing "pull" force comes from the imperfect substitutability between electricity produced from coal and other fuels, $\epsilon \in (1, +\infty)$. Suppose there is no increasing returns to scale. An increase in the relative coal price induces a contemporaneous decrease in relative coal capital investment. This capital imbalance does not persist in subsequent periods. When electricity from coal and other fuels are imperfect substitutes, subsequent periods experience a relative increase in demand for electricity from coal which induces relatively more investment in coal-specific capital. As a consequence, the capital imbalance across the two fuels eventually dissipates.

When both forces are at play, Proposition 2 states that strong path dependence can only be achieved when increasing returns to scale provides a large enough push to overcome the

[25]This comes directly from applying Euler's theorem. Formally, the cross partial derivative of a function $Y(X_{ct}, X_{ct-1})$ of homogeneous degree ψ can be written as

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

Setting $X_{ct} = X_{ct-1}$ so that one can compare the effects of lagged capital investment against current capital investment, $\frac{\partial^2 Y_{ct}}{\partial X_{ct} \partial X_{ct-1}} > -\frac{\partial^2 Y_{ct}}{\partial^2 X_{ct}^2}$ only when $\psi > 1$.

pull from imperfect substitutability, or when $\psi > \frac{-\epsilon}{1-\epsilon}$.[26] In the context of the electricity sector discussed in Section 6.1, this occurs whenever the amplifying force of increasing returns to scale in electricity production offsets the dampening force due to differences in the properties of electricity from different fuels.

7.2 Recovering the elasticity of substitution

The elasticity of substitution between intermediate sectors, ϵ, appears 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). Because these models exhibit different mechanisms for generating path dependence, recovery of ϵ potentially enables this paper's findings to inform settings where another mechanism is driving path dependence in energy transitions. Specifically, I recover ϵ using the mapping between the structural expression in equation (11) and the reduced-form coefficients in equation (4).

To start, denote the structural lagged effect from equation (11) as $\mathcal{B} = \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2}$. The time index in equation (11) is in terms of capital vintages. The time index in my empirical results is in decades relative to the switching event. Noting that the average lifespan of generating units in my sample is 4.7 decades, I convert from event to vintage time by taking the average reduced-form lagged effect across 5 decades. Specifically, using $\hat{\beta}^\tau$ from column 2 of Table 1, the lagged effect of coal price on next-vintage relative coal capital investment is $\widehat{\mathcal{B}} = \frac{1}{5}\sum_{\tau=1}^{5}\hat{\beta}^\tau = -5.17$ with a standard error of 2.83. Together with the estimated scale parameter $\psi = 1.66$, from column 2 of Table 2, the recovered elasticity of substitution is $\epsilon = 4.52$.[27] 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, $\epsilon = 4.52$ falls within the parameter space for which a temporary policy intervention is sufficient to avoid long-term climate disaster, defined as the state of the climate beyond which recovery is not possible.

7.3 Simulating future clean energy transitions

Evidence of strong path dependence implies that it is possible for a temporary relative fuel price shock to induce permanent fuel switching. In the context of climate change policy, it

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

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

$$\epsilon^2[\alpha + \mathcal{B}(\alpha-1)^2] - \epsilon[\alpha + 2\mathcal{B}\alpha(\alpha-1)] + \mathcal{B}\alpha^2 = 0$$

where ϵ is the positive root.

suggests that a temporary policy can induce a long-term decline in carbon emissions. But under what conditions can this occur? In particular, given the abundance of coal resources in the U.S., what is the required magnitude and/or duration of a policy intervention which triggers a sustained clean energy transition?

To analyze how a temporary price shock could overcome low baseline relative coal prices at the national level, I use my theoretical model and estimated structural parameters to simulate future U.S. electricity sector CO_2 emissions following relative coal price shocks of varying magnitude and duration. To ground simulations in recent developments, the magnitude of the shocks considered are benchmarked to recent high relative coal prices following the introduction of natural gas hydraulic fracturing. Figure A.12 plots the ratio of log coal to log natural gas price (both in nominal USD per million BTU) for the U.S. during 1985-2010. As a consequence of hydraulic fracturing, relative coal prices in 2009 and 2010 was 143% higher than what a quadratic trend estimated over 1985-2008 would have predicted.

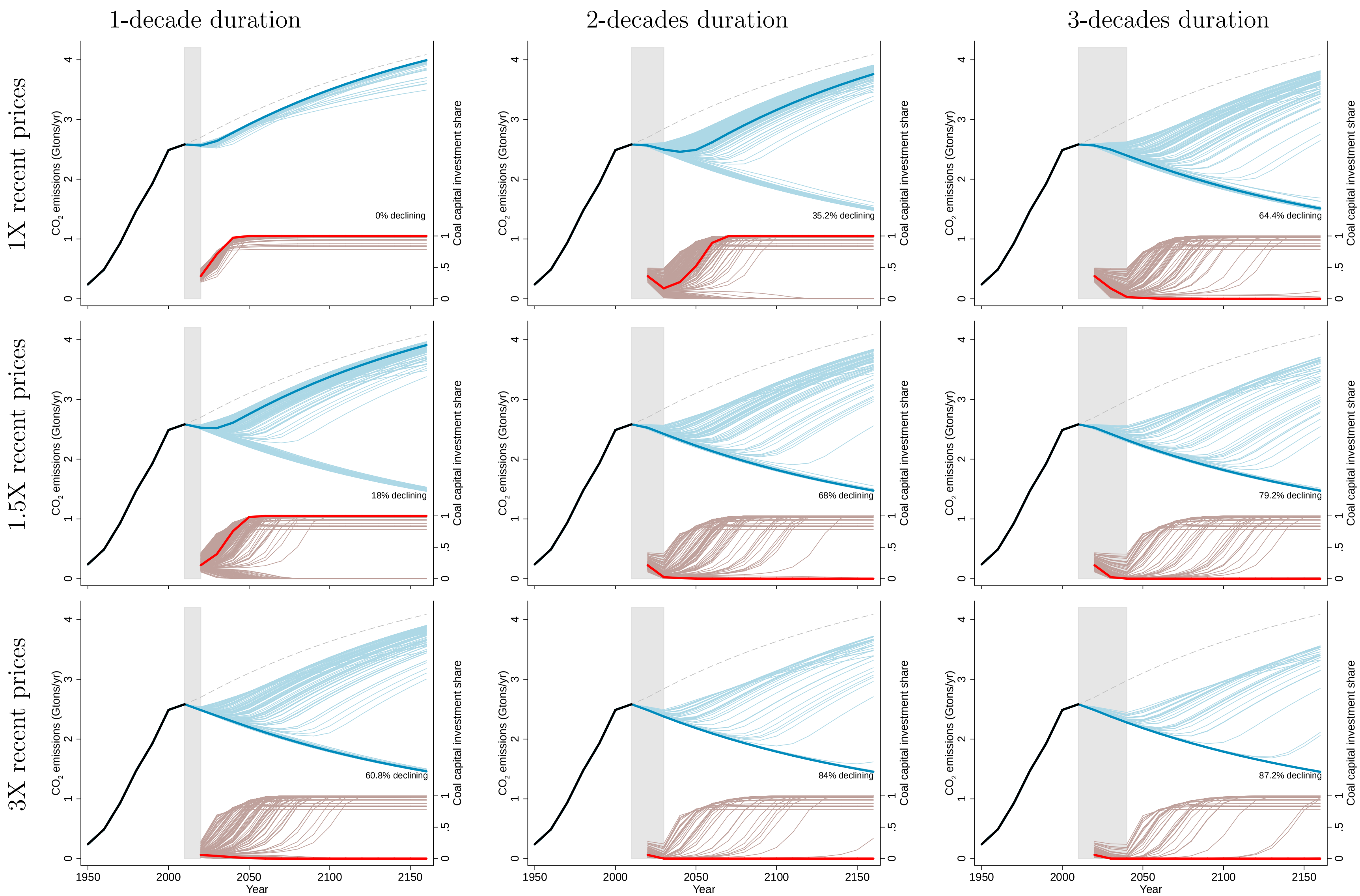
My simulations employ the following simplifying assumptions (see Appendix E for details). First, only electricity from coal and natural gas are considered.[28] Second, to avoid complications from having to forecast future trends in productivity, electricity demand, and other economic conditions, I assume that future productivity, total U.S. coal and natural gas capital, capital depreciation, and baseline relative coal prices in the absence of the shock are all held at recent values. Finally, I assume that the scale and elasticity of substitution parameters are constant throughout the simulation period. Because of these assumptions, it is important to emphasize that these simulations are meant not as actual forecasts of U.S. electricity sector CO_2 emissions but rather as exercises for understanding the conditions under which a permanent transition away from coal could occur.

Figure 5 shows how CO_2 emissions are altered when one varies the magnitude and duration of the temporary relative coal price shock. Top, middle, and bottom panels use price shocks that are 1, 1.5, and 3 times that of recent relative coal prices. Left, middle, and right panels use price shocks that last 1, 2, and 3 decades. Business-as-usual emissions in the absence of the shock is shown as dashed gray lines. When shocks are introduced, the thick solid blue and red lines show CO_2 emissions and the coal share of capital investment under mean estimated values of ψ and ϵ, respectively. The thin solid blue and red lines show CO_2 emissions and coal capital investment share, respectively, from Monte Carlo draws using the estimated distributions of each structural parameter. Each panel also indicates the percentage of draws for which CO_2 emissions are weakly declining in the long-term.[29]

[28] In 2009, coal- and natural gas-fired generating units constituted 92% of total U.S. electricity capital using fossil fuels.

[29] Because natural gas still contains carbon, in none of the simulations do CO_2 emissions reach zero in

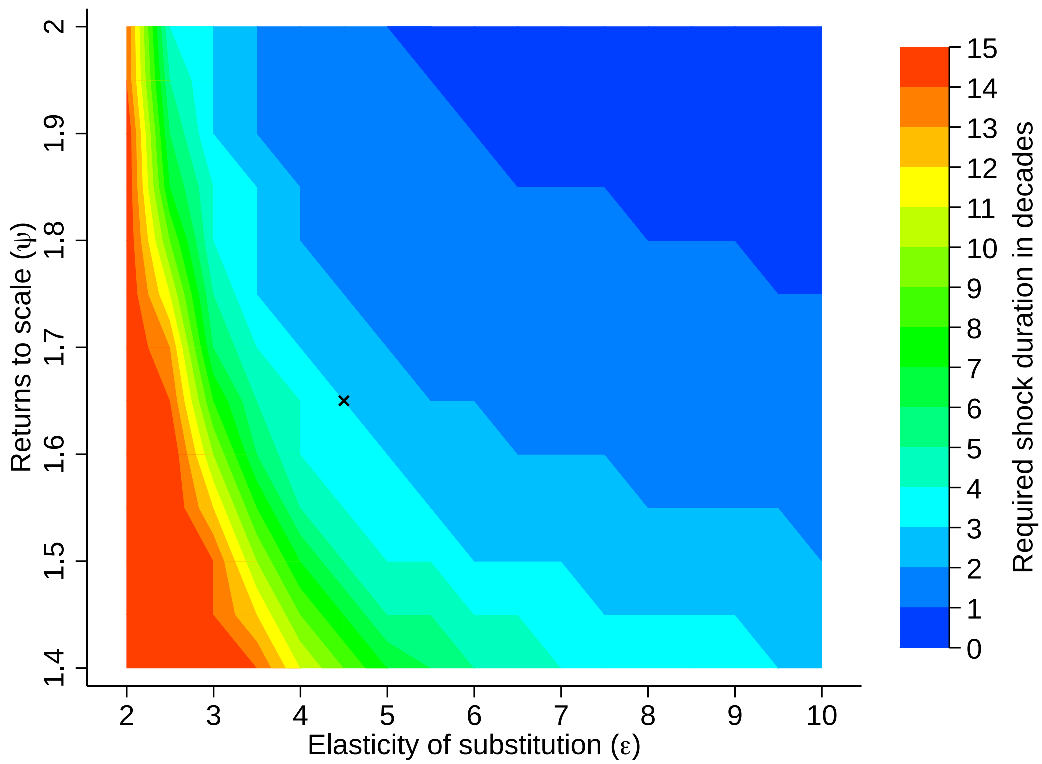
Figure 5: Simulating future U.S. CO_2 emissions following temporary relative coal price shocks



NOTES: Simulations of U.S. CO_2 electricity sector emissions following temporary relative coal price shocks of varying magnitudes and durations. Top, middle, and bottom panels use price shocks 1, 1.5 and 3, times that of recent relative coal prices (see Figure A.12). Left, middle, and right panels use price shocks that last 1, 2, and 3 decades (shown as shaded gray areas). Solid black lines show historical emissions. Dashed gray lines show future business-as-usual emissions without the shock. Thick solid blue and red lines show future projected annual CO_2 emissions (in Gton/yr) and coal capital investment share under the shock, respectively, using mean estimated values of ψ and ϵ. Thin solid blue and red lines show future projected annual CO_2 emissions and coal capital investment share under the shock, respectively, from 250 Monte Carlo draws of the estimated distributions for ψ and ϵ. Percentage of draws with weakly declining long-term CO_2 emissions are also shown. See Appendix E for details.

In general, the likelihood of achieving weakly declining CO_2 emissions in the long-term increases with larger shocks and/or shocks of longer duration. Looking at the top row of Figure 5, if recent relative coal prices were to last for only 1 decade, CO_2 would still rise in the long-term. For a better than 50% chance of weakly declining emissions and a sustained switch away from new investments in coal-fired electricity capital, recent relative coal prices must last at least 3 decades. Going down the first column of Figure 5, if a shock can only last for 1 decade, then it must be at least 3 times that of recent relative coal prices for there to be a greater than 50% chance of achieving weakly declining emissions in the long-term.

Figure 6: Required duration of temporary shock under different structural parameter values



NOTES: Vertical axis shows values for the returns to scale parameter, ψ. Horizontal axis shows values for the elasticity of substitution parameter, ϵ. Heat map shows the minimum number of decades that recent relative coal prices must last in order for long-term CO_2 emissions to be weakly declining. X indicates parameter values estimated in this paper. See Appendix E for details.

Finally, future clean energy transitions may occur under different structural parameters values. Storage technology may improve the intermittency of solar and wind-based electricity, making renewable energy sources better substitutes for coal. Technological advances may uncover new scale economies in electricity production overall. Figure 6 examines future

the long term. Instead, emissions converge asymptotically to a steady-state level where all electricity is produced from natural gas. Thus, the focus of these simulations is on whether temporary policies achieve weakly declining CO_2 emissions in the long-term.

energy transitions under different values of ψ and ϵ. For each pair of parameter values, the heat map in Figure 6 shows the minimum number of decades recent relative coal prices must last in order to achieve weakly declining long-term CO_2 emissions. The require length of the intervention falls when values of ϵ and ψ are higher than those estimated in this paper.

8 Conclusion

This paper estimates path dependence in energy transitions for the U.S. electricity sector over the 20th century. Exploiting shocks to local coal prices driven by the introduction of mechanized mining, I find increasing imbalance in the coal composition of electricity capital lasting ten decades following a shock. Additional evidence detects increasing returns to scale in electricity production as the likely underlying mechanism. Interpreted through a model of the electricity sector with scale-driven structural change, my reduced-form estimates are consistent with a formal definition of strong path dependence whereby temporary shocks of sufficient magnitude and/or duration can induce permanent fuel switching.

This historical evidence is particularly timely given recent developments in the U.S. electricity sector and increasing concerns over climate change impacts. The current spike in relative coal prices following breakthroughs in natural gas extraction is contributing to a slow-down in the construction of new coal-fired power plants. However, these circumstances may not be enough to trigger a sustained transition away from coal in the U.S. electricity sector. Simulations of future energy transitions using this paper's estimated structural parameters demonstrate that a sustained U.S. energy transition away from coal requires either larger or longer lasting interventions.

The presence of strong path dependence in energy transitions also provides a new argument in favor of Pigouvian interventions that directly price an externality over second-best policies that favor specific technologies or alternative resources. It is widely argued that such policies are inefficient when they fail to target cost-effective mitigation strategies. The presence of path dependence amplifies this cost. Suppose a natural gas subsidy induces a transition from coal to natural gas but large climate damages ultimately require even cleaner fuels. Because there is now path dependence in natural gas, the cost of switching once more to a cleaner fuel may be higher now that natural gas has become pervasive. A Pigouvian intervention would avoid this potentially costly detour.

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, and Will Rafey. 2018. "Mirage on the Horizon: Geoengineering and Carbon Taxation Without Commitment." National Bureau of Economic Research Working Paper 24411.

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.

Anderson, Sarah E. 2011. "Complex constituencies: intense environmentalists and representation." *Environmental Politics*, 20(4): 547–565.

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

Atack, Jeremy. 2013. "On the Use of Geographic Information Systems in Economic History: The American Transportation Revolution Revisited." *The Journal of Economic History*, 73(2): 313338.

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

Atkin, David, and Dave Donaldson. 2015. "Who's Getting Globalized? The Size and Implications of Intra-national Trade Costs." National Bureau of Economic Research Working Paper 21439.

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

Averch, Harvey, and Leland L. Johnson. 1962. "Behavior of the Firm Under Regulatory Constraint." *The American Economic Review*, 52(5): 1052–1069.

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. 2018. "Coal Smoke and Mortality in an Early Industrial Economy." *Economic Journal*, forthcoming.

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.

BP. 2017. "Statistical Review of World Energy."

Busse, Meghan R., and Nathaniel O. Keohane. 2007. "Market effects of environmental regulation: coal, railroads, and the 1990 Clean Air Act." *The RAND Journal of Economics*, 38(4): 1159–1179.

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

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

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.

Cicala, Steve. 2015. "When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation." *American Economic Review*, 105(1): 411–44.

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.

Davis, Lucas W., and Catherine Wolfram. 2012. "Deregulation, Consolidation, and Efficiency: Evidence from US Nuclear Power." *American Economic Journal: Applied Economics*, 4(4): 194–225.

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

Donaldson, Dave. 2018. "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure." *American Economic Review*, 108(4-5): 899–934.

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

Energy Information Administration. 1994. "Coal Industry Annual."

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

Fabrizio, Kira R., Nancy L. Rose, and Catherine D. Wolfram. 2007. "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on US Electric Generation Efficiency." *American Economic Review*, 97(4): 1250–1277.

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.

Fujiwara, Thomas, Kyle Meng, and Tom Vogl. 2016. "Habit Formation in Voting: Evidence from Rainy Elections." *American Economic Journal: Applied Economics*, 8(4): 160–88.

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. 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.

Hotelling, Harold. 1931. "The Economics of Exhaustible Resources." *The Journal of Political Economy*, 39(2): 137–135.

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.

Joskow, Paul L., and Nancy L. Rose. 1989. "Chapter 25 The effects of economic regulation." In . Vol. 2 of *Handbook of Industrial Organization*, 1449 – 1506. Elsevier.

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.

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.

Kozhevnikova, Maria, and Ian Lange. 2009. "Determinants of Contract Duration: Further Evidence from Coal-Fired Power Plants." *Review of Industrial Organization*, 34(3): 217–229.

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

MacLeod, W.B., G. Norman, and J.-F. Thisse. 1988. "Price discrimination and equilibrium in monopolistic competition." *International Journal of Industrial Organization*, 6(4): 429 – 446.

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.*

Pigou, Arthur C. 1920. *The Economics of Welfare.* McMillan & Co.

Preonas, Louis. 2018. "Market Power in Coal Shipping and Implications for U.S. Climate Policy." *mimeo.*

Salop, Steven C. 1979. "Monopolistic Competition with Outside Goods." *The Bell Journal of Economics*, 10(1): 141–156.

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.

Tiebout, Charles M. 1956. "A Pure Theory of Local Expenditures." *Journal of Political Economy*, 64(5): 416–424.

Troesken, Werner. 2006. "Regime Change and Corruption. A History of Public Utility Regulation." *Corruption and Reform: Lessons from America's Economic History*, 259–282. University of Chicago Press.

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.

Vogel, Jonathan. 2008. "Spatial Competition with Heterogeneous Firms." *Journal of Political Economy*, 116(3): 423–466.

Vogel, Jonathan. 2011. "Spatial Price Discrimination with Heterogeneous Firms." *The Journal of Industrial Economics*, 59(4): 661–676.

World Bank. 2014. "World Development Indicators."

Appendix to

Estimating Path Dependence in Energy Transitions

FOR ONLINE PUBLICATION

A Data Sources

This section details data used throughout the paper.

A.1 Coal resources, mining, and delivered prices

USGS National Coal Resource Assessment (NCRA) This paper uses two sets of spatial data from the NCRA (East, 2012). The first dataset contains vector-based shape files of Illinois and Appalachian Basin coal resources that are situated less than and greater than 200 feet below the surface. These shape files are used to generate the map in Figure A.3 showing coal resources for the two basins by depth. The second dataset contains characteristics of all coal mines over the Illinois Basin that has operated since 1890. Variables include mine location, opening year, closing year, and area.

Construction of my distance-based shock to delivered coal prices requires the following steps. First, I spatially overlay all large mines in the Illinois Basin that ever existed since 1890 onto shape files of the basin's shallow and deep coal resources, as shown in Figure 3. Using the opening and closing years of each mine, I construct a mine-by-depth-by-decade panel indicating when each shallow or deep coal mine was in operation.[30]

Next, for each county and decade, I search for the nearest mine according to the Euclidean distance between that county's spatial centroid and the mine and note whether it is from a shallow or deep coal resource. This distance is my proxy measure for contemporaneous county delivered coal price, or d_{it} in equation (4).

Finally, for each county, I find the first decade in which a county's nearest mine first switches from extracting shallow to deep coal. Distance to the nearest shallow mine right before the switching event is denoted d_i^0.

FERC-423 forms For each power plant and coal origin county pair, FERC-423 provides annual data on the quantity, price, heat content, sulfur content, and ash content of purchased coal. FERC-423 data is used in three ways. First, for the 1990-1999 period where coal county of origin is more reliable,[31] FERC-423 data on coal heat, sulfur, and heat content is aggregated to the county of origin and averaged across years to produce Table A.1 which documents the heterogeneity in coal quality across the five major U.S. coal basins. Second, data from the 1990-1999 FERC-423 forms are used to test the Herfindahl Principle, as shown in Figure A.7. Third, as verification for my constructed coal prices shocks in Table A.2, I use observed delivered coal prices from the entire set of 1972-1999 FERC-423 forms.

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

[31] According to the EIA, "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... It is very difficult to verify county level data. Users of the data should be aware of this and use the data accordingly."

A.2 Electricity capital and production

EIA-860 forms The EIA-860 forms records the capacity (or capital size), opening and closing years, and primary fuel input at the generating unit level.[32] This paper uses data on generating units from the EIA-860 forms to construct my main outcome variable, relative coal capital at the county-by-decade level covering decades from 1890 to 1990. There are several steps to this construction.

First, I create a cross-sectional dataset of operating and retired generating units, taking the most recent data for each generating unit across the 1990-2012 EIA-860 forms. I then expand this cross-sectional dataset along the time dimension using the opening and closing years of each generating unit to create a generating unit-by-year panel dataset. Next, I sum all generating units that use coal and all generating units that use other fuels to the county-by-decade level. Relative coal capital is defined as the ratio of total capital across generating units that use coal to total capital across generating units that use other fuels. Besides serving as my main outcome variable, relative coal capital is also used to generate Figure 2.

Underlying assumptions behind this data construction procedure are examined in Figures A.8, A.9 and Tables A.3 and A.4, using data from the 1990-2012 EIA-860 forms as well as from the 1980 EIA-860 form, which was digitized for this paper.

Because I know when generating units were built, I can also construct a county-by-decade dataset of relative coal capital investment. Using my generating unit-by-year panel dataset, I sum new generating units that use coal and that use other fuels to the county-by-decade level. Relative coal capital investment is then defined as the ratio of total capital across newly-built generating units that use coal to total capital across newly-built generating units that use other fuels.

A.3 Testing mechanisms

PLATTS/UDI Power plant-level cost data for 1981-1999 was obtained from PLATTS/UDI. It provides non-fuel cost, or the difference between total production costs and fuel costs, which serves as the outcome variable for the return to scale parameter estimates in Table 2.

EIA-923 forms Generating unit-level electricity output and boiler-level coal input data comes from the 2009-2012 EIA-923 forms. Table 3 uses two generating unit-level productivity measures. Following Davis and Wolfram (2012), my capital productivity measure is capital operating performance, or the ratio of a generating unit's electricity output to its capacity. For generating unit g, in power plant p, county i, and state s, 2009-2012 averaged annual capital operating performance is

$$\overline{A}_{Xgpis} = \frac{Y_{gpis}}{X_{gpis} * 8760}$$

[32] I only include generating units owned by public utility companies because units owned by non-utilities are inconsistently reported across EIA-860 forms during the 1990-2012 period.

where Y_{gpis} is annual electricity output (in MWh, from EIA-923 forms), X_{gpis} is capacity (in MW, from EIA-860 forms), and 8760 is the number of hours in a year. My fuel productivity measure is thermal efficiency. For generating unit g, in power plant p, county i, and state s, 2009-2012 averaged annual thermal efficiency is

$$\overline{A}_{Egpis} = \frac{Y_{gpis} * 1000 * 3412}{E_{gpis}}$$

where E_{gpis} is generating unit-level fuel input (in BTU) and 3412 is the equivalent BTU heat content of one KWh of electricity.[33] Generating unit-level E_{gpis} is not directly observed. Instead, the EIA-923 forms provide a boiler-to-generator correspondence, which can have many-to-many matches. To obtain generating unit-level fuel input, I assume that a boiler uniformly divides its fuel input across linked generators. EIA-923 forms prior to 2009 did not include boiler-to-generator correspondences and therefore are excluded from the analysis.

Public Utility Commissions Column 2 of Table A.10 includes only counties in state and decades where there was no Public Utility Commission regulation of electric utilities. The following table summarizes when Public Utility Commission regulation electric utilities was introduced for states in my baseline sample

State	First decade of PUC regulation	Data source
Alabama	1920	Troesken (2006)
Arkansas	1920	State PUC website[34]
Iowa	1970	State PUC website[35]
Illinois	1920	Troesken (2006)
Indiana	1920	Troesken (2006)
Kentucky	1940	State PUC website[36]
Minnesota	1980	State PUC website[37]
Missouri	1920	Troesken (2006)
Mississippi	1960	State PUC website[38]
Tennessee	1920	Troesken (2006)
Wisconsin	1910	Troesken (2006)

[33] See https://www.eia.gov/tools/faqs/faq.cfm?id=107&t=3 for details.
[34] Available: http://www.apscservices.info/commission-history.asp
[35] Available: https://iub.iowa.gov/history
[36] Available: https://psc.ky.gov/Home/About#AbtComm
[37] Available: https://mn.gov/puc/about-us/
[38] Available: https://www.psc.state.ms.us/executive/pdfs/2010/2010%20ANNUAL%20REPORT.pdf

Clean Air Act nonattainment status County-by-year nonattainment status during 1978-1999 under the Clean Air Act comes from the U.S. Environmental Protection Agency. A county-by-decade observation is labeled nonattainment if the county is designated wholly or partially in nonattainment for any of the six criteria pollutant during any year in that decade.[39] This data is used for the estimates in column 4 of Table A.10.

Covariates County-by-decade covariates from 1890 to 1990 come from historical U.S. censuses, collected by Haines (2010).[40] These variables include population, number of manufacturing establishments, and manufacturing employment. To account for changing U.S. county boundaries from 1890 to 1930, I redraw pre-1930 counties to their 1930 spatial definitions to produce a county-by-decade panel of covariates that are spatially consistent over the 20th century. This procedure uses historical GIS county shape files from the U.S. National Historical Geographical Information System (N.H.G.I.S.)[41] and is a modification of the method used by Hornbeck (2012). The resulting covariates serve as controls in column 2 of Table A.6.

Transportation density County-level data on highway and railroad network density in 2010 come from the U.S. Department of Transportation's National Transportation Atlas Database.[42] These variables are used as outcomes in the regressions shown in columns 1 and 2 of Table A.11.

Environmental NGO membership and vote share County-level membership for the Natural Resources Defense Council, The Nature Conservancy, and The National Wildlife Federation in 1996 comes from Anderson (2011). County membership share divides membership by 2000 county population from Haines (2010). County-level Republican Presidential vote share in 2000 comes from Fujiwara, Meng and Vogl (2016). These variables are used as outcomes in the regressions shown in columns 3 and 4 of Table A.11.

A.4 Other

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

U.S. time series data U.S. Census Bureau (1975) provides total and mechanically produced U.S. bituminous coal production during 1890-1950 (shown in Figure A.1) and total electricity

[39] The six regulated criteria pollutants are sulfur dioxide, particulates, nitrogen dioxide, carbon monoxide, ozone, and lead.
[40] Available: http://doi.org/10.3886/ICPSR02896.v3
[41] Available: https://www.nhgis.org{}/documentation/gis-data
[42] Available: http://www.rita.dot.gov/bts/sites/rita.dot.gov.bts/files/publications/national_transportation_atlas_database/2012/index.html

capacity, from fossil fuel and hydropower during 1920-1970 (shown in Figure A.8).[43] Figure A.6 plots the transport cost share of national delivered coal prices during 1902-2007 obtained from McNerney, Farmer and Trancik (2011). Figure A.12 shows national coal and natural gas sales prices during 1985-2011 obtained from the Energy Information Administration (2012).

B Collection and availability of historical data

Sections 4.1 and 4.2 note that the required historical data were either never collected or, if collected, may no longer exist. This section summarizes the data that was historically collected, its relevance for this study, and its known availability today.

B.1 Coal prices

1882-1970

County-level producer coal prices were recorded by the U.S. Geological Survey Bureau of Mine in "Mineral Resources of the United States 1882-1931" and "Mineral Yearbook 1932-1970." This data source, however, does not provide the county-level delivered coal prices needed for this study.

Availability Online.[44]

B.2 Electricity capital

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 at the time by the Department of Commerce and Labor. This survey was repeated in 1907, 1912, and 1917. Unfortunately, this survey classified power plants by prime-mover (i.e. steam, hydro, internal combustion) and not by input fuel, which this paper needs to calculate the fuel composition of electricity capital.

Availability Summaries of these censuses at aggregate data levels are available online.[45] However, power plant-level data could not be located following extensive private conversations with archivists at the National Archives and Records Administration.[46]

[43] Electricity capital data from the U.S. Census Bureau (1975) is broken by steam and hydropower. Steam power typically uses coal, oil, and natural gas as fuel.

[44] Available: hathitrust.org

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

[46] Typically, only 3% of historical government documents are deemed valuable and retained in NARA's permanent collection.

1920-1970:

The Federal Power Commission (FPC), created in 1920, administered annual surveys to document electricity production and fuel consumption. The most important of these were the Annual Financial and Statistical Reports (Form 1) and the Power System Statements (Form 12). Form 1 collects fuel usage at the power-plant level but has two limitations. First, in order to recover fuel-specific capital, fuel-specific capacity factors are needed for each generating unit and are not available. Second, power plant coverage is incomplete. For example, data from the 1948 Form 1 accounts for only 67% of total U.S. steam-powered electricity capital.

Availability Annual state-level statistics for electricity capital by prime-mover and fuel consumption are available in "Production of Energy and Capacity of Plants and Fuel Consumption of Electric Power Plants" as well as in "Electric Power Statistics, 1920-1940".[47] The report titled "Steam-Electric Plant Construction Cost and Annual Production Expenses" has plant-level values from Form 1 and Form 12 for 1948-1974.[48]

1977-1989:

In 1977, the Federal Energy Regulatory Commission (FERC) began publishing the "Inventory of Power Plants in the United States," which combines data on generating units from the Monthly Power Plant Report (Form 4), Annual Power System Statement (Form 12), and the Supplemental Power Statement (Form 12E). This annual inventory includes includes capacity, input fuel, and built year for all operating generating units and those retired each year. Because data on previously retired generating units was not collected, this data cannot be used for reconstructing historical electricity capital.

Availability The 1980 "Inventory of Power Plants in the United States" is available online[49]. It was digitized for the data validation exercises discussed in Section 4.2. Reports from other years are available in microfiche in many research libraries.

1990-:

In 1990, the Energy Information Agency (EIA) began collecting "The Annual Electric Generator Report," (Form EIA-860) which replaced earlier FERC Forms 4, 12, and 12E. EIA notes

> "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

[47] Available: http://hdl.handle.net/2027/mdp.39015023906806

[48] Available: http://catalog.hathitrust.org/Record/000904499

[49] Available: http://hdl.handle.net/2027/umn.31951d02987924n

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 Online.[50]

C Imputing missing small power plants

Power plants with less than 1 MW of combined capacity (or capital) across generating units do not appear in EIA-860 forms. This section explains how missing small power plants are imputed when constructing county-by-decade relative coal capital, $\widetilde{K}_{it} = \frac{K_{cit}}{K_{nit}}$.

For fuel j, county i, decade t, a new power plant indexed by p_{jit} has capacity $X_{p_{jit}}$. Because power plants with $X_{p_{jit}} < 1$ are missing from the EIA-860 data, county total fuel-specific capital investment can be decomposed as

$$\begin{aligned} X_{jit} &= \sum_{p_{jit}:X_{p_{jit}}<1} X_{p_{jit}} + \sum_{p_{jit}:X_{p_{jit}}\geq 1} X_{p_{jit}} \\ &= \underbrace{X_{jit}^{M}}_{\text{Missing}} + \underbrace{X_{jit}^{O}}_{\text{Observed from EIA-860}} \end{aligned}$$

Next, I discretize the support of power plant capacities into 1 MW-wide bins starting at 0.5 MW. Denote f_{jit}^{b} as the number of power plants with capacity $X_{p_{jit}} \in [b-.5, b+.5)$. The missing county total fuel-specific capital investment is then $X_{jit}^{M} = 0.5 * f_{jit}^{0.5}$. My imputation procedure predicts $f_{jit}^{0.5}$. Specifically, to obtain county total fuel-specific capital, K_{jit}, I implement the following procedure for each decade t

1. Estimate $f_t^b = g_t(b) + \text{error}\ \forall b \in \{1.5, \ldots, \bar{b}\}$, where $g_t()$ is a flexible polynomial function

2. Predict $\widehat{f}_t^{0.5} = \widehat{g}_t(0.5)$

3. Downscale national to local capital by using $\widehat{f}_{cit}^{0.5} = \frac{\widehat{f}_t^{0.5} s_t}{N_t}$, where N_t is the number of counties with any operating power plants in decade t and s_t is the national share of electricity capital using coal, aggregated from observed power plants. Likewise, $\widehat{f}_{nit}^{0.5} = \frac{\widehat{f}_t^{0.5}(1-s_t)}{N_t}$. This implies $\widehat{X}_{jit}^{M} = 0.5 * \widehat{f}_{jit}^{0.5}$.

4. Calculate county total fuel-specific capital using $K_{jit} = K_{jit}^{O} + \sum_{\tau=0}(1-\delta)^{\tau}\widehat{X}_{ji,t-\tau}^{M}$, where K_{jit}^{O} is the observed county total fuel-specific capital, $\delta = 0.06$ is the decadal depreciation rate set at 2000s values, and τ is the lagged time index.

[50] Available: http://www.eia.gov/electricity/data/eia860/

Figure A.10 shows the fitted 4th order polynomial function, $\widehat{g}_t()$, estimated for new power plants built in the 1910s (left panel) and in the 1950s (right panel) using observed power plants. The dotted line shows the predicted national frequency of power plants smaller than 1 MW built each decade, or $\widehat{g}_t(0.5)$. Table A.5 shows summary statistics for raw relative coal capital in column 1 and imputed relative coal capital when using a 3rd, 4th, and 5th order polynomial function to fit $g_t()$ in columns 2-4, respectively. Finally, column 5 shows summary statistics for relative coal capital using an alternative, less-informed, imputation for missing small power plants. Specifically, I add 1 MW of coal and non-coal capital investment to each county-by-decade observation. This implicitly assumes that there was a new 1 MW coal-fired power plant and a new 1MW non-coal power plant built in each sample county and decade. I then calculate fuel-specific capital using the capital accumulation expression from step 4 above.

D Theory

D.1 Profit maximization problem

Inputs for older vintaged generating units are fixed. In period t, the power plant chooses capital and fuel inputs for current-vintage generating units. 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 equation (6) and observing that efficient input implies $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ and $A_{Xjt-1}(1-\delta)X_{jt-1} = A_{Ejt-1}E_{jt-1}$ for each vintage- and fuel-specific generating unit, the problem can be written in terms of only capital inputs

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

To explore how scale and productivity effects could generate path dependence for otherwise similar intermediate sectors, 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 equation (A.1) and rewriting in terms of current-vintage relative coal capital investment, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$

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

where $\varphi = (1-\alpha)(1-\epsilon) < 0$. $w_{jt} = \frac{A_{Xjt}}{A_{Ejt}} z_{jt} + r_t$ is the productivity-weighted input price index and $\widetilde{w} = \frac{w_{ct}}{w_{nt}}$ is the relative input price. $\widetilde{A}_{Xt} = \frac{A_{Xct}}{A_{Xnt}}$ is relative capital productivity for generating units of vintage t. Equation (A.2) is equation (7) in the main text.

D.2 Recovering the scale parameter from cost minimization

Consider a power plant p that contains only coal-fired generating units. This allows one to ignore the upper tier of electricity production, drop the fuel index j, and only consider intermediate good production captured by equation (6). Applying efficient allocation for each generating unit, $A_{Xpt}X_{jt} = A_{Ept}E_{pt}$ and $A_{Xpt-1}(1-\delta)X_{pt-1} = A_{Ept-1}E_{pt-1}$, the constrained cost minimization problem can be written in terms of fuel inputs

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

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

$$\min_{E_{pt}} \quad (z_{pt} + \frac{A_{Ept}}{A_{Xpt}} r_t)E_{pt} + z_{pt}Y_{pt}^{1/\alpha}(A_{Ept}A_{Ept-1}E_{pt})^{-1} \tag{A.3}$$

Taking the first order condition of equation (A.3) yields a conditional demand function

$$E^*_{pt} = (Y_{pt})^{1/2\alpha}(\frac{z_{pt}}{z_{pt} + \frac{A_{Ept}}{A_{Xpt}}})^{1/2}(A_{Ept}A_{Ept-1})^{-1/2} \tag{A.4}$$

Inserting equation (A.4) into non-fuel cost at the cost-minimizing input level, $\text{non\_fuel\_cost}_{pt} = C(z_{pt}, r_{pt}, Y_{pt}) - z_{pt}(E^*_{pt} + E_{pt-1}) = r_t(\frac{A_{Ept}}{A_{Xpt}} E^*_{pt})$, and applying a log transformation

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

where $\psi = 2\alpha$. Equation (A.5) is the structural analog to the OLS specification in equation (8) from the main text with one exception. Equation (A.5) is missing the cost of labor which, for ease of exposition, is not included as a factor of production in the intermediate good production function shown in equation (6). Empirically, labor costs will be a component of the error term in equation (8).

E Simulating future emissions

This section details the procedure for simulating future CO_2 emissions following a relative coal price shock, as shown in Figure 5.

E.1 Parameters

- Lagged effect of coal price on next-vintage relative coal capital investment: $\widehat{\mathcal{B}} = \frac{1}{5}\sum_{\tau=1}^{5} \hat{\beta}^{\tau} = -5.17$, with standard error $\widehat{\sigma}_{\mathcal{B}} = 2.83$ (from column 2 of Table 1).
- Returns to scale: $\widehat{\psi} = 1.66$, with standard error $\widehat{\sigma}_{\psi} = 0.24$ (from column 2 of Table 2).
- Baseline relative coal prices: $\widetilde{w}_t^o = 0.4$ (based on value in 2000s)
- Relative coal price shock: $\Delta = 1.43$ (based on Figure A.12).
- Relative productivity: $\widetilde{A}_{Xt} = 0.7$ (based on value in 2000s).
- Capital depreciation rate: $\delta = 0.06$ (based on value in 2000s).
- Carbon content of coal: $C_c = 4931.3$ lb CO_2/short ton coal.[51]
- Carbon content of natural gas: $C_n = 119.9$ lb CO_2/thousand cubic feet.[52]

E.2 Historical emissions

For each decade $t \in [1950, 2000]$, U.S. electricity sector CO_2 emissions is

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

where E_{ct} and E_{nt} is coal (in short tons) and natural gas (in thousand cubic feet) consumed respectively by the U.S. electricity sector in decade t.[53]

E.3 Simulating future emissions

For each combination of shock duration, $d \in \{10, 20, 30\}$, and shock multiplier, $M \in \{1, 1.5, 3\}$, define the time series of relative coal prices as

$$\widetilde{w}_t = \widetilde{w}_t^o + \Delta * M * \mathbf{1}(t \leq 2000 + d)$$

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

[51] Available here: https://www.eia.gov/environment/emissions/co2_vol_mass.cfm
[52] Available here: https://www.eia.gov/environment/emissions/co2_vol_mass.cfm
[53] E_{ct} and E_{nt} obtained from the Energy Information Administration (2012).

- Draw $\psi(b) \sim N(\widehat{\psi}, \widehat{\sigma_\psi})$ and $\mathcal{B}(b) \sim N(\widehat{\mathcal{B}}, \widehat{\sigma}_{\mathcal{B}})$. Define $\alpha(b) = \frac{\psi(b)}{2}$
- Obtain $\epsilon(b)$ by taking the positive root of[54]

$$\epsilon(b)^2[\alpha(b) + \mathcal{B}(b)(\alpha(b)-1)^2] - \epsilon(b)[\alpha(b) + 2\mathcal{B}(b)\alpha(b)(\alpha(b)-1)] + \mathcal{B}(b)\alpha(b)^2 = 0$$

  Define $\varphi(b) = (1-\alpha(b))(1-\epsilon(b))$

- For each future decade $t \in [2010, 2150]$

  1. Apply equation 11 to obtain relative coal capital investment:

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

  2. Obtain coal capital investment while holding total capital fixed

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

  3. Obtain natural gas capital investment while holding total capital fixed

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

  4. Obtain coal capital

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

  5. Obtain natural gas capital

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

  6. Obtain total CO_2 emissions using 2000 emissions intensity

$$M(b)_t = K(b)_{ct}\frac{E_{c2000}}{K_{c2000}}C_c + = K(b)_{nt}\frac{E_{n2000}}{K_{n2000}}C_n$$

Figure 5 plots CO_2 emissions, $M(b)_t$, and the coal capital investment share, $\frac{X(b)_{ct}}{X(b)_{ct}+X(b)_{nt}}$, for the 250 Monte Carlo draws across each combination of shock duration, $d \in \{10, 20, 30\}$, and shock

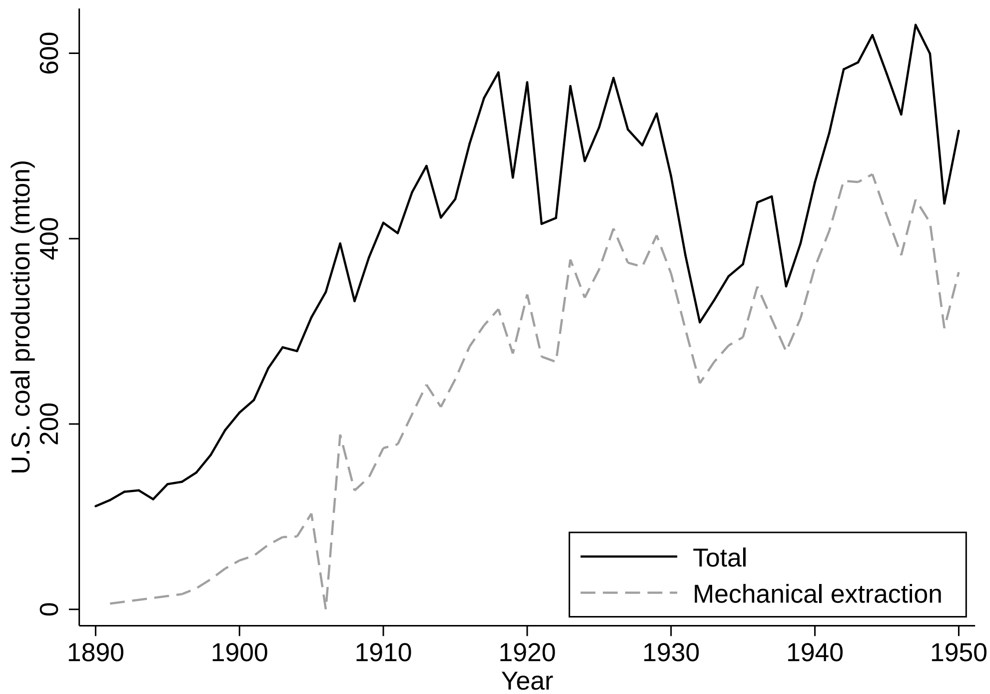
[54] For given parameters, there are occasional cases when either the solutions to this quadratic formula are imaginary numbers or the positive root is $\epsilon < 1$. My simulation discards these draws, which occurred 28 times out of a total of 278 draws.

multiplier, $M \in \{1, 1.5, 3\}$. It also shows the percentage of Monte Carlo draws for which long-term CO_2 emissions are weakly declining.

To explore the consequences of different structural parameter values, Figure 6 replicates the above simulation procedure for different fixed values of ϵ and ψ with the shock multiplier set at $M = 1$. The heat map plotted in Figure 6 shows the minimum shock duration d needed in order for long-term CO_2 emissions to be weakly declining.

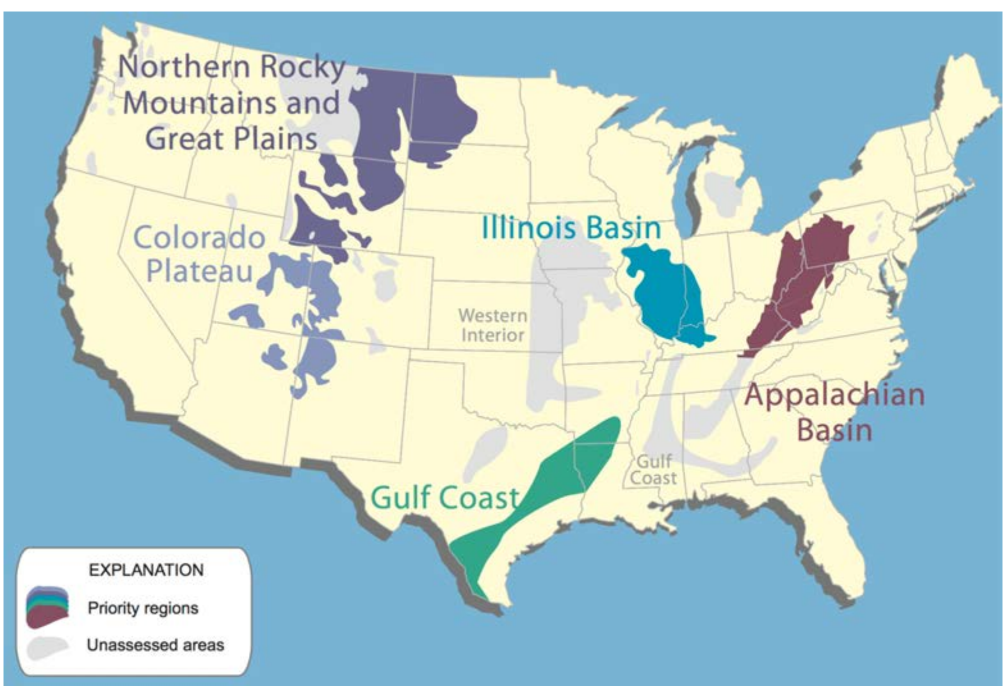
Appendix Figures

Figure A.1: U.S. bituminous coal production and mechanization



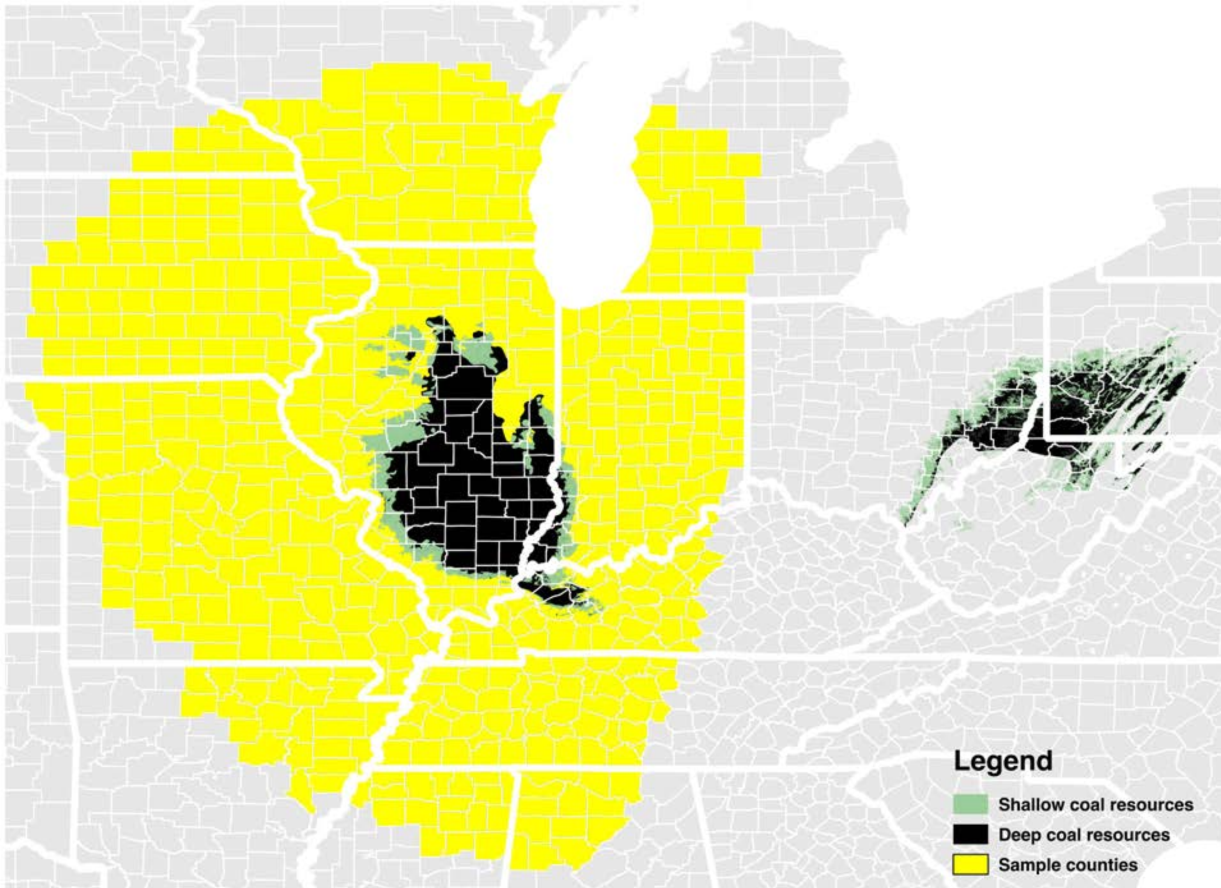
Notes: Annual U.S. bituminous coal production (in mega short tons). Solid black line shows coal production. Dashed gray line shows coal production from mechanical extraction. Data from U.S. Census Bureau (1975).

Figure A.2: U.S. coal basins



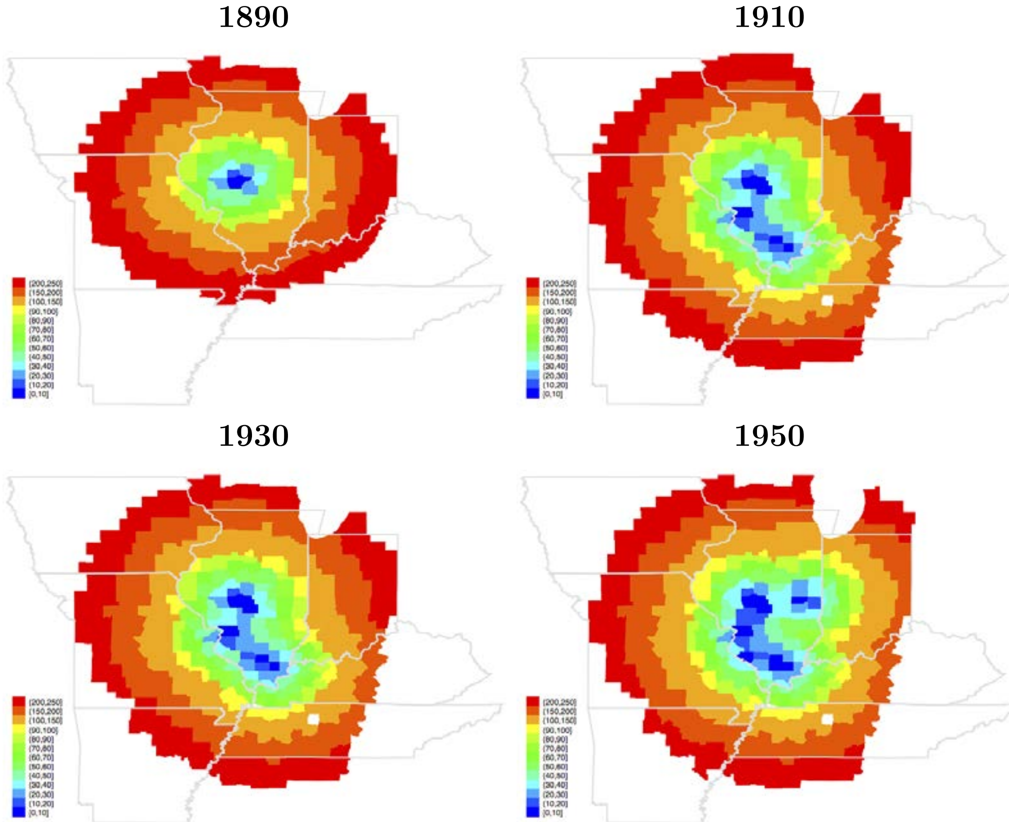
Notes: Map of major U.S. coal basins. Reproduced from East (2012).

Figure A.3: Location of sample counties and coal basins



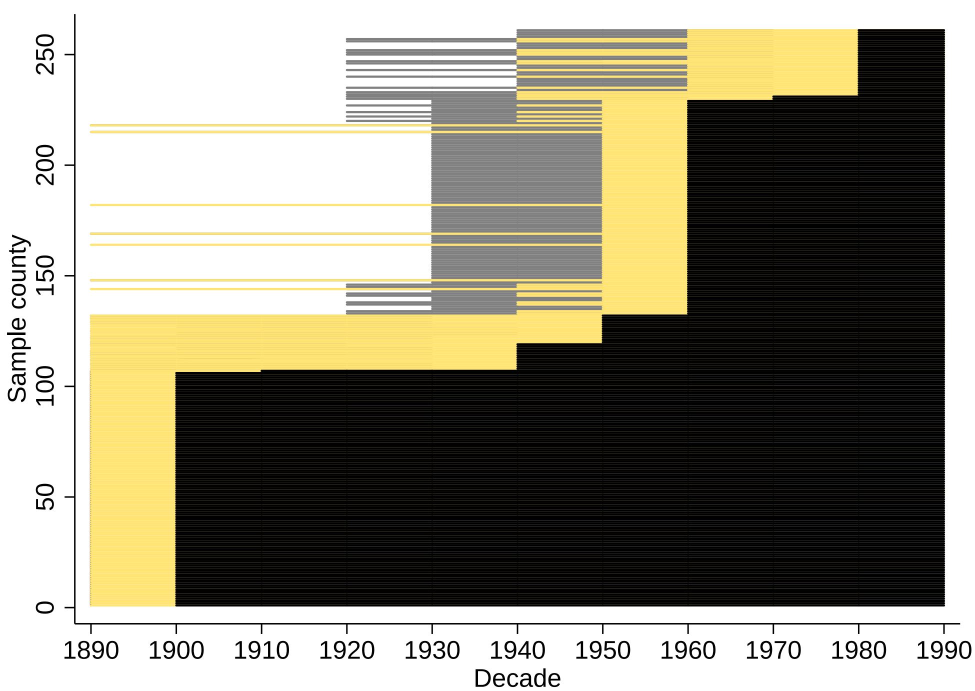
NOTES: A county is included in the baseline sample (in yellow shading) if its spatial centroid is (i) closer to coal resources in the Illinois Basin than in the Appalachian Basin and (ii) less than 250 miles from nearest Illinois coal resource. Shallow (< 200 ft. underground) and deep (> 200 ft. underground) Illinois and Appalachian Basin coal resources also shown in green and black shading, respectively.

Figure A.4: County distance to nearest mine by decade



NOTES: County distance to nearest coal mine during 1890-1950 over sample counties.

Figure A.5: Timing of shallow to deep coal switching for each sample county

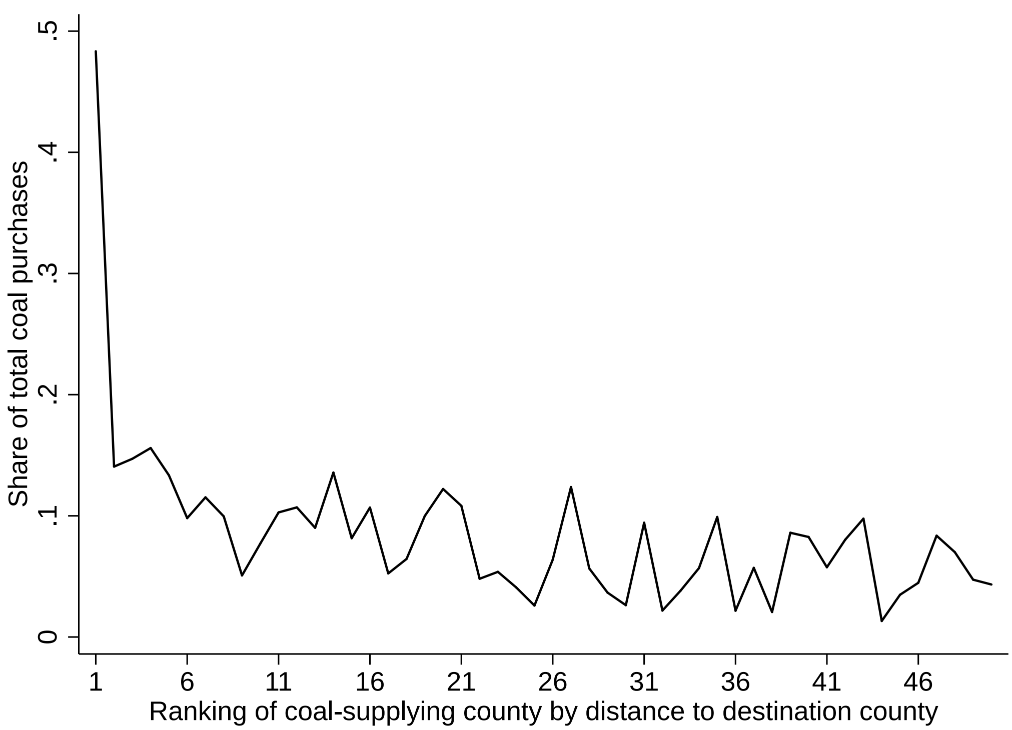


NOTES: The timing of when a sample county's nearest mine switches from a shallow to coal mine for the first time. Counties are stacked according to the decade when the switching event occurs. The gray, yellow, and black shaded areas correspond to event-time periods $h < 0$, $h = 0$, and $h > 0$, respectively.

Figure A.6: Share of delivered coal price due to transport costs at the national level

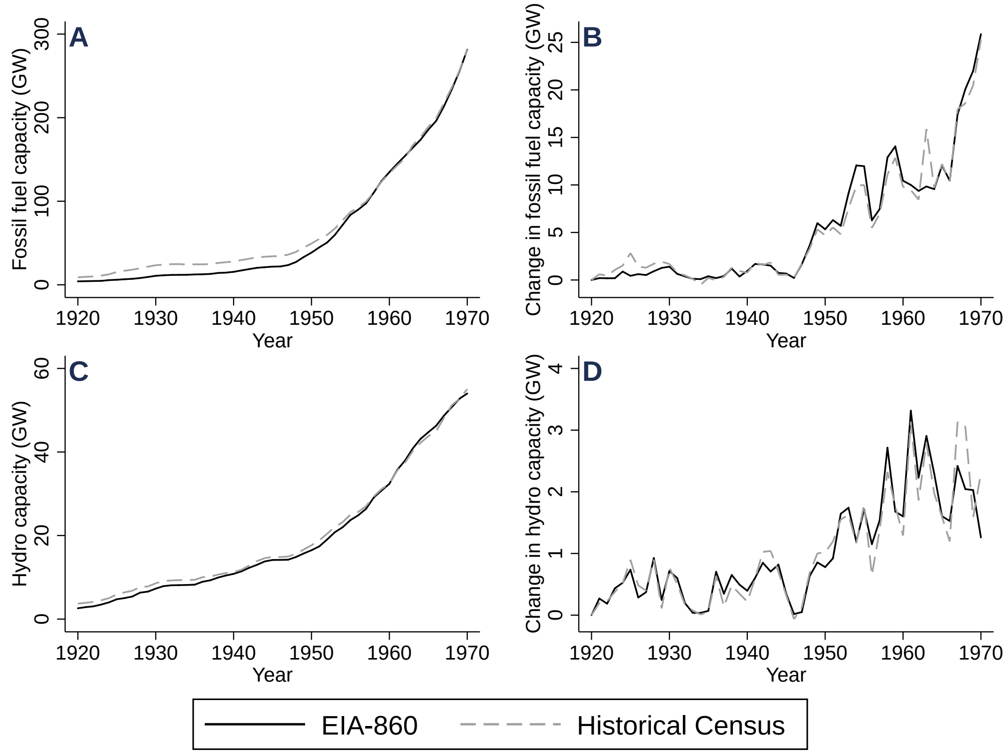
NOTES: Share of transport cost (nominal USD per short ton) in delivered coal price (nominal USD per short ton) for the U.S. during 1902-2007. Data from McNerney, Farmer and Trancik (2011).

Figure A.7: Testing the Herfindahl Principle



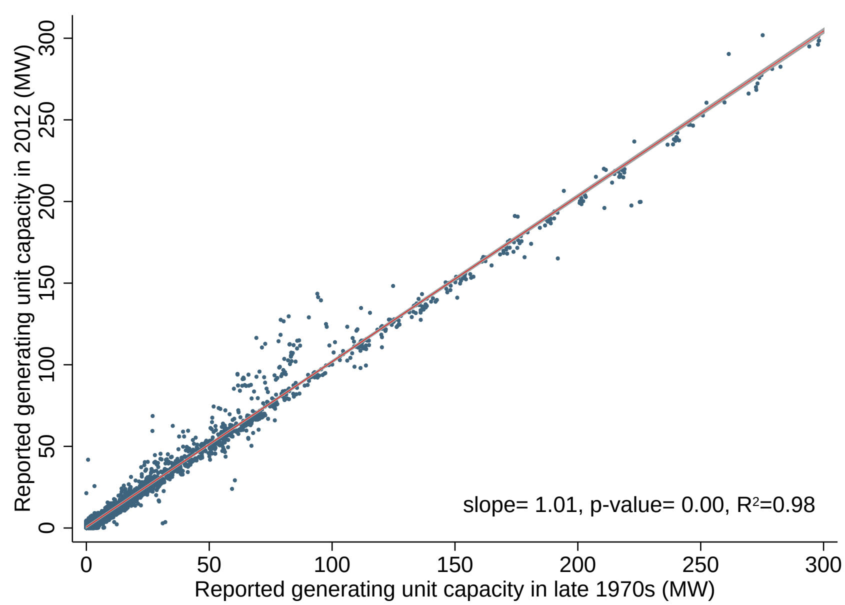
NOTES: Vertical axis shows share of total coal purchase from destination county. Horizontal axis shows the ranking of bilateral distance between spatial centroids of origin and destination counties. Data averaged over 1990-1999 and all U.S. counties that purchase coal for electricity.

Figure A.8: Comparing U.S. electricity capacity from EIA-860 vs. historical census



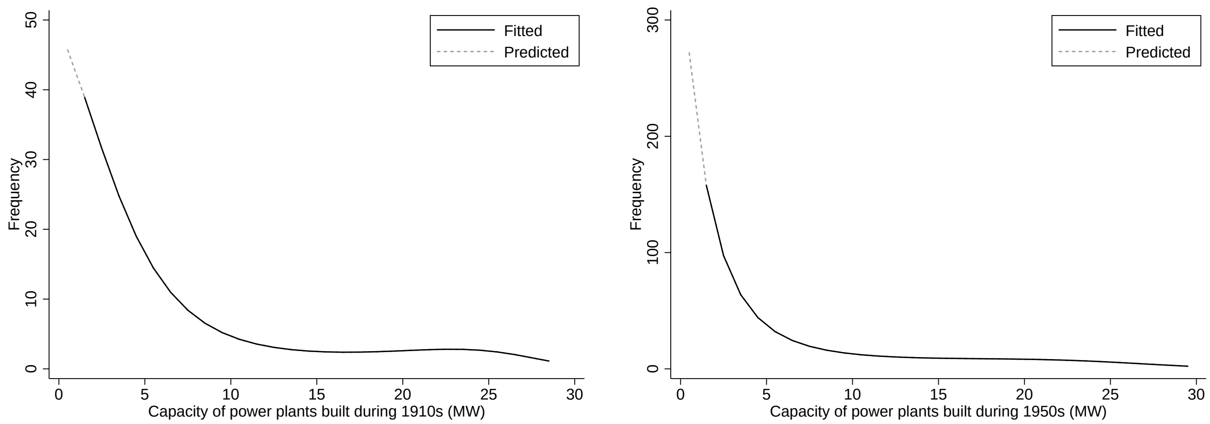
Notes: Panel (A) compares annual aggregate U.S. electricity capacity (in GW) using fossil fuels constructed from EIA-860 forms (solid black lines) against values from the U.S. Historical Census (dashed gray lines) over 1920-1970. Panel (B) plots capacity changes. Panels (C) and (D) show the same information as in panels (A) and (B) but for aggregate U.S. hydropower capacity.

Figure A.9: Comparing generating unit capacity in late 1970s and 2012



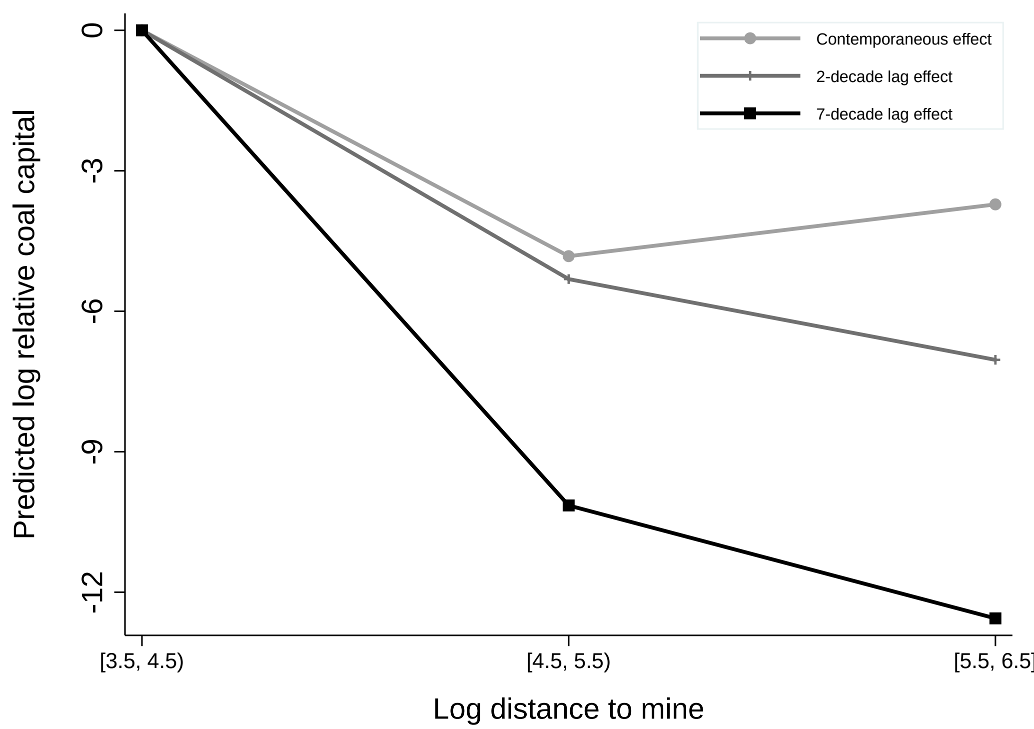
Notes: Scatter shows reported generating unit capacity (in MW) in the 2012 EIA-860 form against reported generating unit capacity in the late 1970s (in MW). Both axes are truncated at 300 MW. Linear fit in red with 95% confidence interval shown in gray. Coefficient and p-value shown from a linear regression using all matched generating units.

Figure A.10: Fitted and predicted capacity distribution of power plants by built decade



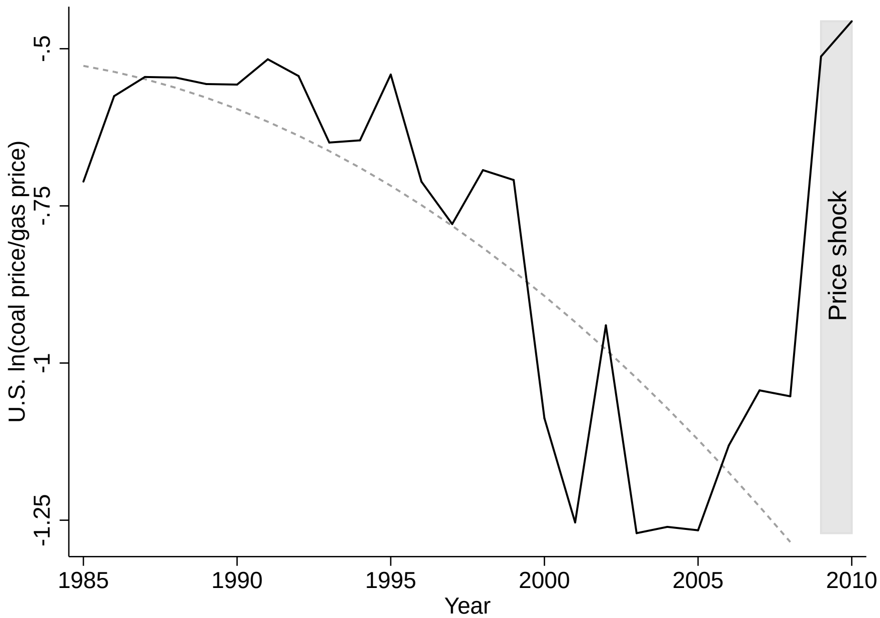
Notes: Solid line shows fitted capital frequency distribution of all power plants built during 1910s (left panel) and 1950s (right panel). Fitted relationship uses a 4th order polynomial function for power plants with capacity less than 30 MW and greater than 1 MW. Dashed line shows predicted capacity frequency for power plants with capacity less than 1 MW. See Appendix C for details.

Figure A.11: Testing for nonlinearity in the relative coal capital-distance relationship



NOTES: Plot examines whether log relative coal capital is linear in log distance to the nearest mine using a version of equation (4) with discretized log distance bins. Log relative coal capital is predicted using log distance to contemporaneous nearest mine, and log distance to shallow mine two and seven decades after the switching event. Each log distance variable is broken into 1 log distance wide bins. For each period, the predicted log relative coal capital is normalized to the value of the omitted log distance bin, defined as log distances between 0 and 3.5.

Figure A.12: Ratio of U.S. coal to natural gas sales price



Notes: Solid black line shows log ratio of U.S. coal sales price to U.S. natural gas sales price (both in nominal USD per million BTUs) during 1985-2010. Dashed gray line shows quadratic time trend estimated over 1985-2008.

Appendix Tables

Table A.1: Coal quality heterogeneity by basin

Coal basin	No. of counties	Std. dev. in heat content	Std. dev. in ash content	Std. dev. in sulfur content
Appalachian	187	657.24 (1)	2.96 (3)	0.93 (5)
Colorado	79	1159.77 (3)	4.47 (5)	0.15 (1)
Gulf	56	1370.36 (5)	3.71 (4)	0.29 (3)
Illinois	279	807.30 (2)	2.52 (2)	0.84 (4)
Northern Rockies	65	1270.79 (4)	1.79 (1)	0.22 (2)

NOTES: Standard deviation in coal heat, ash, and sulfur content across counties that produce coal in each coal basin. Basin ranking for each characteristic in parentheses (1=least heterogeneous). County-level values calculated using 1990-1999 averages.

Table A.2: Observed delivered coal prices vs. constructed distance-based measure

	(1)	(2)	(3)
	Outcome is log delivered coal price		
$\ln d_{it}$	0.381**	0.582***	0.342***
	(0.150)	(0.153)	(0.113)
Decade	1970s	1980s	1990s
Counties	153	153	133

NOTES: Each column is a separate cross-sectional regression of observed log delivered coal price (in nominal USD per ton) averaged within each decade on log distance to nearest mine and state fixed effects. County sample shown in Figure A.3. Columns (1), (2), and (3) use data from the 1970s, 1980s, and 1990s, respectively. Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.3: Comparing generating unit characteristics across 1990-2012 EIA-860 forms

	Percentage of generating units with different reported values			
Number of different values	Capacity	Primary fuel	Opening year	Retirement year
0	74.78	94.25	96.88	80.01
1	2.62	1.49	0.52	1.98
2	1.81	0.57	0.44	2.92
3	1.07	0.23	0.08	0.87
4	1.26	0.2	0.08	0.87
5	0.85	0.51	0.19	1.5
6	0.69	0.46	0.24	0.67
7	0.95	0.29	0.29	0.72
8	0.53	0.27	0.22	1.08
9	0.79	0.2	0.14	1.74
10	0.72	0.25	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 unit-level characteristics. Each cell shows the percentage of 1990-2011 EIA-860 forms with a reported value that is different from that reported in the 2012 EIA-860 form. For example, row 1, column 1 indicates that 76.78% of generating units reported the same capacity in 1990-2011 as was reported in 2012.

Table A.4: Comparing generating unit primary fuel in late 1970s and 2012

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. For example, 92.2% of generating units which reported coal as the primary fuel in the 1970s also reported coal in 2012.

Table A.5: Summary statistics for raw and imputed relative coal capital

	(1) Raw	(2) Imputed 3rd order poly.	(3) Imputed 4th order poly.	(4) Imputed 5th order poly.	(5) Imputed add 1 MW
			Number of observations		
Total	2,371	2,371	2,371	2,371	2,371
Missing	1,248	0	0	0	0
Zero	825	0	0	0	0
Other	298	2,371	2,371	2,371	2,371
			Summary statistics		
Obs	1,123	2,371	2,371	2,371	2,371
Mean	9.71	64.95	75.42	65.08	8.53
Median	0	.15	.16	.14	1
SD	53.44	355.17	416.33	356.31	37.28
Skewness	11.51	7.39	7.48	7.41	6.76

Notes: Top panel of each column shows the number of total, missing, and zero-value, and other observations for my county-by-decade estimating sample. Bottom panel shows various summary statistics. Column 1 shows the raw unadjusted relative coal capital. Columns 2-4 adds imputed missing power plants with capacity less than 1 MW by using 3rd, 4th, and 5th order polynomial functions for $g_t()$, respectively (see Appendix C). Column 5 adjusts the raw variable by adding 1 MW to both coal and non-coal capital investment to every observation.

Table A.6: Robustness: additional identification concerns

	(1)	(2)	(3)	(4)
	Outome is relative coal capital			
$\ln d_i^0$ (β^τ)				
2 decades lead	-1.38	-0.69	-0.91	-1.49
	(1.02)	(1.61)	(1.07)	(0.96)
1 decade lead	-0.66	-0.22	-1.45	-1.14
	(0.67)	(0.85)	(0.95)	(0.71)
	–	–	–	–
1 decade lag	-0.68	-0.72	-0.60	-2.22**
	(1.25)	(1.05)	(1.29)	(1.01)
2 decades lag	-4.11***	-3.19***	-4.95***	-4.92***
	(1.15)	(1.07)	(1.25)	(1.39)
3 decades lag	-3.75***	-4.11***	-3.83***	-4.29***
	(0.66)	(1.39)	(0.76)	(0.62)
4 decades lag	-3.51***	-4.11***	-2.87***	-4.33***
	(0.71)	(1.53)	(0.88)	(0.67)
5 decades lag	-4.58***	-5.96***	-3.31***	-5.40***
	(0.98)	(1.68)	(0.98)	(1.09)
6 decades lag	-3.71***	-4.84***	-2.85***	-4.53***
	(0.75)	(1.43)	(0.91)	(0.67)
7 decades lag	-6.20***	-6.99***	-3.30**	-7.12***
	(1.37)	(1.81)	(1.67)	(1.45)
8 decades lag	-7.30***	-8.04***	-4.36**	-8.27***
	(1.59)	(1.93)	(1.77)	(1.71)
9 decades lag	-7.30***	-8.02***	-4.62**	-8.22***
	(1.56)	(2.05)	(1.96)	(1.69)
10 decades lag	-7.01***	-7.95***	-4.27**	-7.86***
	(1.56)	(2.05)	(1.89)	(1.68)
$\ln d_{it}$ (π)	-1.53***	-1.56**	-2.20**	-1.69***
	(0.53)	(0.61)	(0.86)	(0.55)
County covariates	No	Yes	No	No
Mine fixed effect	No	No	Yes	No
Dropped if shallow mine becomes 2nd nearest	No	No	No	Yes
Observations	2,371	2,108	2,371	2,035
Counties	261	261	261	219

NOTES: Estimates of β^τ and π from equation (4) using a Poisson model. Outcome variable is relative coal capital at the county-by-decade level. County sample shown in Figure A.3. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates. Column 2 augments baseline model with county-by-decade population, number of manufacturing establishments, and manufacturing employment, all in logs. Column 3 augments baseline model with shallow mine fixed effects. Column 4 estimates baseline model but drops any county for which the shallow mine becomes the second nearest mine in any decade after the switching event. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.7: Robustness: imputing missing small power plants

	(1)	(2)	(3)	(4)	(5)
	Outcome is relative coal capital				
$\ln d_i^0$ (β^τ)					
2 decades lead	-1.38	-1.31	-1.32	-0.23	-2.66
	(1.02)	(1.01)	(1.00)	(0.46)	(2.22)
1 decade lead	-0.66	-0.70	-0.69	0.0069	-1.53
	(0.67)	(0.67)	(0.67)	(0.31)	(0.94)
	–	–	–	–	–
1 decade lag	-0.68	-0.58	-0.59	-0.12	-0.84
	(1.25)	(1.19)	(1.19)	(0.33)	(0.58)
2 decades lag	-4.11***	-3.97***	-3.97***	-1.90***	-2.49
	(1.15)	(1.07)	(1.08)	(0.44)	(2.02)
3 decades lag	-3.75***	-3.70***	-3.68***	-2.11***	-4.18**
	(0.66)	(0.66)	(0.66)	(0.41)	(2.03)
4 decades lag	-3.51***	-3.44***	-3.43***	-2.03***	-4.97**
	(0.71)	(0.71)	(0.71)	(0.37)	(1.94)
5 decades lag	-4.58***	-4.49***	-4.46***	-2.29***	-3.91*
	(0.98)	(0.91)	(0.91)	(0.37)	(2.08)
6 decades lag	-3.71***	-3.65***	-3.63***	-1.47***	-2.82
	(0.75)	(0.75)	(0.74)	(0.48)	(2.20)
7 decades lag	-6.20***	-6.15***	-6.12***	-3.18***	0.71
	(1.37)	(1.36)	(1.36)	(0.80)	(2.96)
8 decades lag	-7.30***	-7.24***	-7.22***	-3.90***	-3.33
	(1.59)	(1.58)	(1.57)	(0.80)	(2.64)
9 decades lag	-7.30***	-7.25***	-7.22***	-3.95***	-2.63
	(1.56)	(1.56)	(1.55)	(0.78)	(2.45)
10 decades lag	-7.01***	-6.96***	-6.94***	-3.76***	-2.20
	(1.56)	(1.55)	(1.54)	(0.75)	(2.83)
$\ln d_{it}$ (π)	-1.53***	-1.54***	-1.54***	-0.86**	-2.41***
	(0.53)	(0.53)	(0.53)	(0.40)	(0.75)
Small plant imputation	4^{th} ord. poly	3^{rd} ord. poly	5^{th} ord. poly	add 1MW	none
Observations	2,371	2,371	2,371	2,371	382
Counties	261	261	261	261	66

NOTES: Estimates of β^τ and π from equation (4) using a Poisson model. Outcome variable is relative coal capital at the county-by-decade level. County sample shown in Figure A.3. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates which imputes missing small power plants with a 4th order polynomial function for $g_t()$ to construct construct relative coal capital (see Appendix C). Column 2 uses a 3rd order polynomial function for $g_t()$. Column 3 uses a 5th order polynomial function for $g_t()$. Column 4 adds 1 MW to both coal and non-coal relative capital investment for each observation to construct relative coal capital. Column 5 uses the raw relative coal capital variable. Of the 1,123 observations that are non-missing, column 5 drops 724 observations from counties where the outcome variable is always zero and 17 observations from counties with only one non-missing variable per county to estimate a Poisson fixed effects model. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Robustness: county sample

	(1)	(2)	(3)	(4)
	Outcome is relative coal capital			
$\ln d_i^0$ (β^τ)				
2 decades lead	-1.38	-2.22*	-1.21	-1.41
	(1.02)	(1.27)	(1.02)	(1.05)
1 decade lead	-0.66	-1.27	-0.49	-0.76
	(0.67)	(0.82)	(0.75)	(0.66)
	–	–	–	–
1 decade lag	-0.68	-1.99	-0.24	-0.71
	(1.25)	(1.23)	(1.17)	(1.05)
2 decades lag	-4.11***	-6.01***	-2.71***	-3.94***
	(1.15)	(1.16)	(1.03)	(1.03)
3 decades lag	-3.75***	-4.61***	-2.34***	-3.70***
	(0.66)	(0.83)	(0.90)	(0.65)
4 decades lag	-3.51***	-4.50***	-2.01**	-3.49***
	(0.71)	(0.93)	(0.96)	(0.69)
5 decades lag	-4.58***	-7.31***	-1.69	-4.14***
	(0.98)	(2.07)	(1.12)	(0.77)
6 decades lag	-3.71***	-4.49***	-1.59	-3.84***
	(0.75)	(0.98)	(1.16)	(0.73)
7 decades lag	-6.20***	-6.91***	-3.71**	-4.56***
	(1.37)	(1.36)	(1.50)	(0.97)
8 decades lag	-7.30***	-8.09***	-4.62***	-5.16***
	(1.59)	(1.54)	(1.69)	(1.21)
9 decades lag	-7.30***	-8.10***	-4.43***	-5.12***
	(1.56)	(1.48)	(1.66)	(1.23)
10 decades lag	-7.01***	-7.76***	-4.18**	-5.06***
	(1.56)	(1.48)	(1.67)	(1.23)
$\ln d_{it}$ (π)	-1.53***	-1.53***	-1.33**	-0.74
	(0.53)	(0.52)	(0.52)	(0.59)
County sample	Baseline	<200 miles from Ill. basin	<300 miles from Ill. basin	Include closer to App. basin
Observations	2,371	1,938	2,881	3,218
Counties	261	208	320	338

NOTES: Estimates of β^τ and π from equation (4) using a Poisson model. Outcome variable is relative coal capital at the county-by-decade level. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 uses baseline county sample shown in Figure A.3. 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. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Robustness: alternative modeling choices

	(1)	(2)	(3)
$\ln d_i^0$ (β^τ)			
2 decades lead	-1.38	-0.27	0.32
	(1.02)	(0.42)	(1.04)
1 decade lead	-0.66	0.27	0.27
	(0.67)	(0.40)	(0.64)
	–	–	–
1 decade lag	-0.68	-0.42	-0.76*
	(1.25)	(0.26)	(0.39)
2 decades lag	-4.11***	-1.21***	-1.40**
	(1.15)	(0.31)	(0.60)
3 decades lag	-3.75***	-1.71***	-1.93**
	(0.66)	(0.49)	(0.87)
4 decades lag	-3.51***	-1.80***	-2.36**
	(0.71)	(0.55)	(1.04)
5 decades lag	-4.58***	-2.18***	-2.61**
	(0.98)	(0.66)	(1.06)
6 decades lag	-3.71***	-1.79**	-2.49**
	(0.75)	(0.85)	(1.12)
7 decades lag	-6.20***	-2.89**	-3.23
	(1.37)	(1.29)	(2.04)
8 decades lag	-7.30***	-4.10***	-4.20**
	(1.59)	(1.37)	(2.02)
9 decades lag	-7.30***	-3.71**	-3.86*
	(1.56)	(1.51)	(2.04)
10 decades lag	-7.01***	-3.74**	-3.55*
	(1.56)	(1.71)	(2.08)
$\ln d_{it}$ (π)	-1.53***	-0.18	-0.66
	(0.53)	(0.62)	(0.90)
Model	Poisson	Linear	Neg. bin.
Observations	2,371	2,371	2,371
Counties	261	261	261

NOTES: Estimates of β^τ and π from equation (4). Outcome variable is at the county-by-decade level. County sample shown in Figure A.3. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates using a Poisson model with relative coal capital as the outcome. Column 2 uses a log-log linear model with log relative coal capital as the outcome. Column 3 uses a negative binomial model with dispersion parameter that is a function of the expected mean of relative coal capital, the outcome. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Mechanism: cost-of-service and Clean Air Act regulations

	(1)	(2)	(3)	(4)
	Outome is relative coal capital			
$\ln d_i^0$ (β^τ)				
2 decades lead	-1.38	-0.75	-1.61	0.093
	(1.02)	(0.91)	(1.00)	(1.39)
1 decade lead	-0.66	-0.71	-1.46	0.53
	(0.67)	(0.56)	(1.00)	(1.12)
	–	–	–	–
1 decade lag	-0.68	-2.80**	-1.91**	-3.09**
	(1.25)	(1.09)	(0.83)	(1.39)
2 decades lag	-4.11***	-2.93***	0.032	-3.26**
	(1.15)	(1.11)	(0.87)	(1.29)
3 decades lag	-3.75***	-2.93*	-3.12***	-4.14***
	(0.66)	(1.55)	(1.07)	(0.80)
4 decades lag	-3.51***	-6.99*	-2.46	-3.70***
	(0.71)	(3.92)	(2.02)	(1.13)
5 decades lag	-4.58***	-3.11***	-5.56*	-2.24*
	(0.98)	(1.13)	(2.91)	(1.26)
6 decades lag	-3.71***	-3.12***	-4.50	2.11
	(0.75)	(1.14)	(2.95)	(2.19)
7 decades lag	-6.20***		-4.85*	-4.30*
	(1.37)		(2.57)	(2.25)
8 decades lag	-7.30***			-5.75**
	(1.59)			(2.42)
9 decades lag	-7.30***			-5.94***
	(1.56)			(2.14)
10 decades lag	-7.01***			-5.31**
	(1.56)			(2.10)
$\ln d_{it}$ (π)	-1.53***	0.36	0.75	-1.76***
	(0.53)	(1.11)	(1.13)	(0.67)
Drop PUC	No	Yes	No	No
Drop ever in nonattainment	No	No	No	Yes
Sample period	1890-1990	1890-1970	1890-1960	1890-1990
Observations	2,371	746	1,588	1,685
Counties	261	201	261	185

NOTES: Estimates of β^τ and π from equation (4) using a Poisson model. Outcome variable is relative coal capital at the county-by-decade level. County sample shown in Figure A.3. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates. Column 2 drop county-decade observations when there is a state Public Utility Commission regulating electric utilities. Column 3 includes only observations during 1890-1960. Column 4 drops counties that were ever designated as nonattainment under the Clean Air Act. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Mechanism: upstream and downstream sectors

	(1)	(2)	(3)	(4)
	Outcome is			
	ln railroad density	ln highway density	env. NGO share	republican vote share
$\ln d_i^o \times sinceEvent_i$ (ω_1)	-0.048	-0.022	-0.000022	0.0033
	(0.031)	(0.018)	(0.000078)	(0.0034)
$\ln d_i^o$ (ω_2)	-0.0019	0.013	-0.00057	-0.0022
	(0.17)	(0.11)	(0.00046)	(0.021)
$sinceEvent_i$ (ω_3)	0.23*	0.051	-0.000095	-0.011
	(0.14)	(0.072)	(0.00031)	(0.014)
Counties	458	458	458	458

NOTES: Estimates from equation 10 using county-level outcomes. All models includes state and NERC region fixed effects, and county centroid longitude and latitude. County sample shown in Figure A.3. Outcome in column 1 is log railroad density in 2010 (in miles per square mile). Outcome in column 2 is log highway density in 2010 (in miles per square mile). Outcome in column 3 is the population share of individuals who are members of three major environmental NGOs in 1996 (in %). Outcome in column 4 is the share of eligible voters who voted for the Republican Presidential candidate in 2000 (in %). Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.