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 December 2025

I thank Lint Barrage, Dan Benjamin, Javier Birchenall, Hoyt Bleakley, Chris Costello, Clément de Chaisemartin, Melissa Dell, Jonathan Dingel, Walker Hanlon, Kelsey Jack, Akshaya Jha, Per Krusell, David Lagakos, Derek Lemoine, Gary Libecap, Suresh Naidu, Trevor O'Grady, Ryan Oprea, 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. Jason Benedict, Kayleigh Campbell Bierman, Danae Hernandez Cortes, and Simon Jean provided excellent research assistance. I declare that I have no relevant or material financial interest related to this paper, nor any conflicts of interest. The views expressed herein are those of the author and do not necessarily reflect the views of the National Bureau of Economic Research.

Estimating Path Dependence in Energy Transitions
Kyle C. Meng
NBER Working Paper No. 22536
August 2016, Revised December 2025
JEL No. N51, N52, O41, Q35, Q43, Q54, Q58

ABSTRACT

What induces clean energy transitions? When transitional dynamics exhibit strong path dependence, a temporary shock to input composition can trigger permanent structural change. I examine whether such dynamics characterize the U.S. electricity sector's use of coal - the most climate-damaging fuel - across the 20th century. Exploiting local coal supply shocks driven by changing regional accessibility of subsurface coal, I find increasing imbalance in the coal composition of electricity capital lasting ten decades following a shock. A structural change model enables recovery of a key substitution parameter from reduced-form estimates and explores conditions for triggering sustained future clean energy transitions.

Kyle C. Meng
University of California, Santa Barbara
Department of Economics
and NBER
kmeng@bren.ucsb.edu

1 Introduction

Economies tend to consume more abundant resources first, turning to alternatives only after sufficient depletion has set in (Herfindahl, 1967). However, if consumption of an abundant resource generates substantial social costs, an earlier transition to alternative resources may be required. Many environmental challenges are characterized by this problem, and none more so than anthropogenic climate change. Carbon emissions arise, in part, because the most abundant fossil fuel, coal, is also the most climate-damaging.[1] As a consequence, it is widely recognized that the global economy must permanently transition away from using coal in order to address climate change.[2]

How can a sustained energy transition away from coal be induced? Economic theory offers two perspectives. In the traditional view, where an economy's composition of resources is determined primarily by relative supply, a *permanent* intervention that lowers coal use (e.g., a Pigouvian tax) is needed to offset coal's supply advantage. If the policy were ever removed, the forces of relative supply would return, enabling coal consumption and carbon emissions to resume upward trajectories. This behavior appears in economies exhibiting either no or weak path dependence in energy transitions. Unfortunately, permanent policy interventions may be unrealistic when governments have difficulty committing to long-term policies. Indeed, the history of climate policies to date is filled with examples of policy revisions, reversals, and withdrawals.[3]

In contrast, recent structural change models posit that in the presence of certain transitional dynamics, a large but *temporary* intervention that exogenously lowers coal use can permanently overcome coal's abundant supply (Acemoglu et al., 2012, 2016; Lemoine, 2024; Fried, 2018; Acemoglu et al., 2023). Under such circumstances, a sustained long-term transition away from coal could be achieved even after the intervention is lifted. Economies with this feature are broadly characterized as having strong path dependence in energy transitions. Whether such dynamics actually govern energy transitions, however, remains an open

[1]Coal contains over half of energy stored in global fossil fuel deposits (BP, 2017). It is also responsible for over half of emitted anthropogenic carbon dioxide since the pre-industrial era (Boden, Marland and Andres, 2013).

[2]Local pollution from coal inspired early papers on external costs (Pigou, 1920; Coase, 1960), and continues to motivate an extensive valuation literature (e.g., Chay and Greenstone (2003, 2005); Barreca, Clay and Tarr (2014); Clay, Lewis and Severnini (2024); Beach and Hanlon (2018); Hanlon (2020)).

[3]For example, the U.S. recently withdrew from the U.N. Paris Agreement and repealed clean energy subsidies under the 2022 Inflation Reduction Act. 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 (2023) for other examples.

empirical question.

This paper provides one of the first empirical studies of long-run energy transitions. Specifically, I examine the transitional dynamics of the U.S. electricity sector over the 20th century, making three contributions. First, I find reduced-form evidence of strong path dependence in long-run transitions between coal and other fuels, shedding light on the historical circumstances that made the U.S. among the most carbon-intensive economies in the world. Second, I combine additional evidence with a model of structural change for the electricity sector to recover the long-run elasticity of substitution between coal and other fuels from reduced-form estimates, an important common parameter found across recent optimal climate policy models but which remains largely unknown to date. Third, I conduct simulations using my calibrated model to characterize how climate policies of varying magnitudes and durations could enable a permanent future U.S. energy transition away from coal.

There are two main empirical challenges in estimating energy transitions in the electricity sector. First, electricity capital (i.e., power plants) is built for specific fuels and lasts multiple decades. Detecting significant changes in fuel composition therefore requires sufficiently long data series spanning multiple capital vintages. Second, estimating path dependence requires an exogenous and temporary shock to coal supply that subsequently alters the fuel composition of electricity capital.

To overcome the first challenge, I combine modern and historical power plant records to construct a new dataset of county-level, fuel-specific electricity capital for the U.S. midwest across the 20th century. This enables an analysis of long-run energy transitions capturing changes in the fuel composition of electricity capital. To address the identification challenge, I construct local coal supply shocks using local coal transport distances driven by the changing regional accessibility of subsurface coal. The introduction of mechanized mining around the early 20th century allowed extraction over previously inaccessible coal held in deep underground deposits. Mechanized mining, together with the location and subsurface depth of coal resources, altered coal transport distances and thus the spatial distribution of delivered coal prices. As such, these *local* coal transport distance shocks are driven primarily by two remote *regional* factors - the time-invariant spatial structure of subsurface coal geology and time-varying mining technology - and thus plausibly uncorrelated with unobserved local determinants of fuel composition.

Using an event study design with county-by-decade panel data, I find evidence of strong path dependence: a negative coal supply shock triggers a declining trajectory of the relative use of coal in electricity capital that lasts for up to ten decades. Notably, these lagged effects display discrete jumps at two and seven decades after the event, corresponding to the expected timing of two subsequent vintages of electricity capital. In support of my parallel trends assumption, I do not find pre-trends in key covariates or in the outcome variable. A series of robustness checks addresses further identification, data construction, sample restriction, and statistical modeling concerns.

Path dependence in energy transitions emerges from a combination of "push" and "pull" forces. Prior literature highlights various mechanisms that amplify energy transitions. These include macroeconomic channels such as directed technical change (Aghion and Howitt, 1992; Acemoglu et al., 2012, 2016; Aghion et al., 2016; Fouquet, 2016; Hassler, Krusell and Olovsson, 2021; Lemoine, 2024; Casey, 2024) and network externalities (David, 1985; Aghion et al., 2025). They also include microeconomic channels such as increasing returns to scale and local productivity effects such as via learning-by-doing that may be more relevant with my county-level research design (Nerlove, 1963; Christensen and Greene, 1976; Arthur, 1994). The literature also considers a common dampening force against runaway transitions: imperfect long-run substitutability between clean and dirty fuels, a key structural parameter found across several recent optimal climate policy models (Acemoglu et al., 2012; Golosov et al., 2014; Lemoine, 2024; Fried, 2018; Acemoglu et al., 2023). In general, the need for reliable estimates of energy input substitutability has been a long-standing concern in the energy economics literature.[4] Most prior estimates use price variation at the annual or sub-annual levels (Lanzi and Wing, 2010; Aghion et al., 2016; Papageorgiou, Saam and Schulte, 2017; Knittel, Metaxoglou and Trindade, 2019; Jo, 2025), which may not capture important long-run patterns of substitution involving changes in energy capital.[5]

To recover a substitution parameter relevant for my empirical context, I first develop a model of structural change for the electricity sector at the county level featuring increasing returns to scale and local productivity effects. Empirical tests using the structure of this model verify that increasing returns to scale may be driving my reduced-form estimates. Further tests fail to statistically detect alternative explanations related to local productivity effects as well as other possible mechanisms such as cost-of-service electricity regulation, the U.S. Clean Air Act, coal procurement contracts, increasing returns in coal transportation,

[4]See Atkinson and Halvorsen (1976); Griffin and Gregory (1976); Pindyck (1979); Stern (2012); Papageorgiou, Saam and Schulte (2017); Fried (2018); Casey (2024).

[5]A related paper by Hawkins-Pierot and Wagner (2025) shows that negative energy price shocks at the time a manufacturing facility opens results in persistently higher energy efficiency during the lifetime of that facility, regardless of subsequent energy prices. This technological lock-in effect has similar climate policy implications as the fuel composition dynamics explored in this paper by creating path dependence in energy use.

and residential household sorting, though statistical uncertainty prohibits one to definitively rule out these mechanisms. This evidence, together with the structure of the model enables a mapping between my reduced-form estimate of strong path dependence and the long-run elasticity of substitution between coal and other fuels.

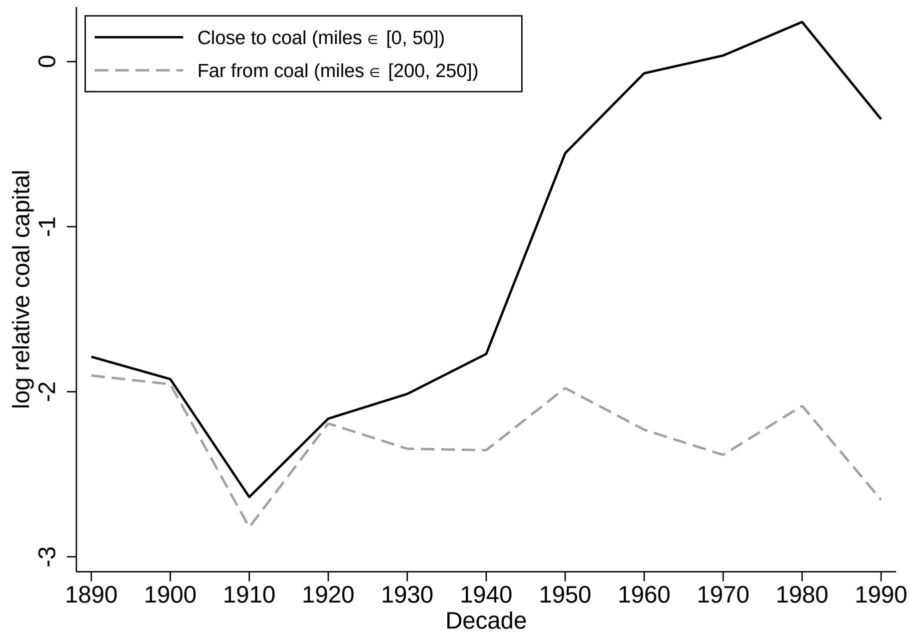
Finally, I calibrate my model using reduced-form estimates to draw lessons for future U.S. energy transitions away from coal. Evidence of strong path dependence implies it is possible for a temporary intervention to induce permanent fuel switching. But under what conditions? To answer this question, I simulate future electricity sector carbon emissions for the average U.S. county following temporary relative coal price shocks of varying magnitude and duration. For a better than 50% chance of achieving a permanent switch away from coal and thus weakly declining carbon emissions, a temporary shock equal in magnitude to recent high relative coal prices (e.g., due to natural gas hydraulic fracturing) must last at least five decades. Alternatively, if the shock can only last one decade, it must then be six times higher than that of recent prices to trigger sustained fuel switching. Further simulations explore how requirements for sustained energy transitions change under different elasticity of substitution and scale parameter values. Altogether, these simulations conclude that in the absence of climate policy, recent economic conditions are insufficient for sustaining a permanent U.S. energy transition away from coal.

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 the long-run elasticity of substitution, and simulate future carbon emissions. Section 8 concludes.

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

The United States has one of the most carbon-intensive economies in the world. Figure A.1 plots carbon dioxide (CO_2) emissions per capita against GDP per capita for non-OPEC countries in 2000 using data from Boden, Marland and Andres (2013) and Bank (2014). U.S. emissions per capita is nearly 2 standard deviations higher than what income would predict. This reflects the U.S. electricity sector's heavy reliance on coal, the most carbon-intensive of energy fuels.[6] Since the 1960s, roughly 40% of U.S. electricity has been produced from coal (Energy Information Administration, 2012). Why is the U.S. electricity sector so dependent on coal? Many observers point to its world-leading coal resources. However, a casual exploration of historical patterns of local coal use suggests that a supply-based explanation may be incomplete.

Figure 1: 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.

Before turning to these patterns, it is useful to first introduce how coal composition is measured throughout this paper. In the electricity sector, capital size is usually denoted by the capacity of a generating unit, or the maximum electricity it can produce in an hour. Thus, a natural measure of coal composition is the ratio of the capacity of coal-fired generating units to the capacity of generating units using other fuels. I call this relative coal capital.[7]

[6] 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.

[7] Structural change concerns changes in the composition of inputs, and not that of individual inputs. Because inputs are jointly determined, it is important to examine an outcome variable that is a function of both coal and non-coal electricity capital. Relative coal capital is preferred over coal capital share (i.e., the ratio of capacity of coal-fired generating units to total capacity of generating units across all fuels), as my main outcome variable, largely because predictions from standard models of structural change (discussed further in Sections 6 and 7) are typically expressed as factor input ratios.

Figure 1 examines U.S. relative coal capital at the county-by-decade level (see Section 4.2 for data construction details).[8] 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 studied in this paper (for reasons discussed in Section 4.1). Consistent with a coal supply argument, counties closer to coal resources exhibit higher relative coal capital throughout the 20th century. However, the gap between these two sets of counties widens dramatically over this period. Such a pattern cannot be explained by coal supply alone, which predicts relative coal capital convergence, not divergence, over time as counties closer to coal resources face higher mine prices as their supplying mines deplete more rapidly. Instead, these divergent dynamics hint at the presence of path dependence, under which past relative coal capital has a direct effect on future relative coal capital.

Unfortunately, Figure 1 does not provide causal evidence of path dependence in energy transitions. Among other concerns, the dynamics displayed in Figure 1 may reflect the role of unobserved time-invariant and varying characteristics with ongoing influence on relative coal capital, rather than the direct influence of past relative coal capital. The next section details this and other statistical considerations.

3 Empirical framework

This section begins by describing the empirical challenges to identifying path dependence in energy transitions. I then discuss how shocks to local coal transport distance driven by the changing regional accessibility of subsurface coal may overcome these challenges. To focus for now on empirical issues, this section considers path dependence and its strength within a reduced-form setting. Section 7 will offer formal definitions through the lens of a structural change model.

3.1 Challenges to identifying path dependence

Statistically, path dependence is present when past outcomes have a causal effect on future outcomes. Path dependence in structural change occurs when the outcome exhibiting such dynamics is the composition of inputs. I begin with a simple empirical framework to illustrate the challenges with estimating path dependence in energy transitions. There are two decades,

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

$t \in \{1, 2\}$, 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 for each period is

$$\ln \widetilde{K}_{i1} = \pi \ln w_{ci1} + \xi_{i1} \tag{1a}$$

$$\ln \widetilde{K}_{i2} = \rho \ln \widetilde{K}_{i1} + \pi \ln w_{ci2} + \xi_{i2} \tag{1b}$$

where π is the contemporaneous price effect. ρ is my reduced-form parameter for path dependence. The error term ξ_{it} contains unobserved time-varying and -invariant determinants of relative coal capacity. The presence of the latter, in particular, implies that a component of ξ_{i2} also appears in $\widetilde{K}_{i1}$ such that the autoregressive coefficient obtained by directly estimating equation (1b) may not distinguish path dependence following a transitory shock from the persistent effects of time-invariant determinants.[9] To formalize this concern, insert equation (1a) into (1b)

$$\ln \widetilde{K}_{i2} = \rho\pi \ln w_{ci1} + \pi \ln w_{ci2} + \varepsilon_{i2} \tag{1c}$$

where $\varepsilon_{i2} = \rho\xi_{i1} + \xi_{i2}$. Two statistical assumptions are needed for the ratio of lagged to contemporaneous effects to identify ρ. The first is $E[w_{ci2}\varepsilon_{i2}|w_{ci1}] = 0$, which states that contemporaneous coal prices must be exogenous. This exogeneity assumption would be violated if, for example, relative coal capital and prices were simultaneously determined. The second identifying assumption is $E[w_{ci1}\varepsilon_{i2}|w_{ci2}] = 0$, which requires that past coal prices be uncorrelated with unobserved contemporaneous determinants of relative coal capital. If this "exclusion restriction" assumption is satisfied, past prices affect current relative capital only through past relative capital. When both assumptions are satisfied, lagged effects that are larger in magnitude than contemporaneous effects suggest strong path dependence in energy transitions. Conversely, lagged effects that are smaller in magnitude than contemporaneous effects suggest weak path dependence.

In practice, another complication arises when estimating equation (1c), which implicitly assumes that lagged effects can be detected within a single decade. Electricity capital decays slowly over multiple decades. To detect effects on subsequent new capital investments (and not just on depreciated existing capital), 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, are no longer available today (see Appendix B for a

[9] See a related discussion in Bleakley and Lin (2012).

summary of historical data collection and availability).

3.2 Solution: regionally-driven local coal transport distances

To address these empirical challenges, I construct shocks to local coal transport distance that rely on plausibly weaker identifying assumptions and span a 110-year period. The basic idea is to construct *local* shocks using the changing *regional* accessibility of subsurface coal.

Mechanized mining and access to deep coal resources Prior to the 20th century, most coal in the U.S. was manually mined which generally limited extraction to coal resources 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. As shown in Figure A.2, nearly the entire production increase in bituminous coal - the variety most used for electricity - between 1890-1930 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. The interaction between this aggregate technology shock and the spatial distribution of deep coal resources altered local delivered coal prices.

Using local coal transport distances To illustrate how changing regional coal accessibility from the introduction of mechanized mining can be exploited, I return to the empirical framework in Section 3.1. Suppose only shallow coal resources no deeper than 200 feet from the surface could be extracted in the first decade. In the second decade, technology enables deep coal mining. Denote the set of operating mines in each period as M_t.

To construct local coal transport distance shocks in this setting, I enlist the Herfindahl Principle (Herfindahl, 1967), a theoretical result in 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. While the Herfindahl Principle is a theoretical result, later in the paper I examine its empirical validity (see Figure A.8) and conduct robustness checks in case its underlying assumptions are violated (see Section 5.3).[10]

[10]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 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 $m_t \in M_t$. Delivered coal price can then be decomposed into

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

where ζ_{it} includes other supply-side factors. The coefficient on $\ln d_{it}$ is normalized to one for reasons detailed below.

Inserting equation (2) into equation (1c) yields

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

where $\mu_{i2} = \rho\pi\zeta_{i1} + \pi\zeta_{i2} + \varepsilon_{i2}$. $\ln d_{i2}$ 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.[11] To ensure $\frac{\partial ln\widetilde{K}_{i2}/\partial ln d_{i1}}{\partial ln\widetilde{K}_{i2}/\partial ln d_{i2}}$ 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. Observe that the coefficient on distance in equation (2) need not be one. As long as this coefficient is time-invariant[12] and the identifying assumptions for equation (3) are satisfied, $\frac{\partial ln\widetilde{K}_{i2}/\partial ln d_{i1}}{\partial ln\widetilde{K}_{i2}/\partial ln d_{i2}}$ would cancel out this coefficient and identify ρ. That is, to estimate path dependence, one only needs the identifying variation - in this case local coal transport distances - to drive relative coal capital. It need not directly capture coal prices, per se.

Compared with directly using coal prices in equation (1c), use of $\ln d_{it}$ has two distinct advantages. First, construction of $\ln d_{it}$ requires only historical data on the location, coal depth, and operating years of coal mines, which, unlike historical coal prices, is available across the 20th century. Second, observe that outside of county i's location (addressed below), $\ln d_{it}$ is driven primarily by two *regional* determinants: (i) the time-invariant spatial distribution of the depth of subsurface coal resources and (ii) the time-varying introduction of mechanized mining which made deep coal accessible. As a consequence, use of $\ln d_{it}$ is

for a generalized setting with multiple, spatially differentiated, consumers. Robustness checks in Section 5.3 considers potential complications that arise when coal resources have heterogeneous quality and when coal mines have market power.

[11] 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.

[12] Table A.2 shows there is no trend in the county-level coal price-transport distance correlation from the 1970s to the 1990s, the period in which both variables are observed.

more likely to satisfying my identifying assumptions.

3.3 Regression specification

Equation (3) from Section 3.1's simple empirical framework is a cross-sectional regression. In practice, I use a county-by-decade panel dataset over the 20th century to estimate an event study specification that generalizes equation (3). This approach affords several additional advantages. First, identifying assumptions are now in terms of parallel trends, with county-specific fixed effects removing the potentially confounding influence of time-invariant factors such as geography and other natural features. Namely, in the absence of shocks to local coal transport distance, relative coal capital would have followed the same trends in all sample counties. Second, because each county experiences the event of first switching to deep coal at different moments in time, lagged effects of distance to the nearest shallow mine in "event time" can be identified relative to local secular trends through the use of state-by-decade fixed effects. Third, the presence of multiple decades of data for each county enables estimation of lead effects and lagged effects over a long time horizon. Lead effects provide an indirect test of parallel trends. Lagged effects across multiple decades allow sufficient time for capturing new electricity capital decisions.

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.[13] $\ln d_i^0$ is distance to that shallow coal mine and represents my local coal transport distance shock. $h < 0$ indicates an earlier period when a county's coal supplier was yet a different shallow coal mine. Using this event study setting, I estimate for county i, in state s, during decade t

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

where γ^τ captures common event-time effects and μ_{ist} is an error term. County fixed effects, ϖ_i, removes concerns that local coal transport distance shocks may be correlated with a

[13]Note that in our baseline specification, $h = 0$ can span multiple decades if the shallow coal mine is a county's nearest supplier for more than a decade. In a robustness check, we redefine $h = 0$ as the single last decade before a county first switches to deep coal to examine whether there are differential pre-trends in the decades before that switch.

county's time-invariant characteristics such as its geography.[14] State-by-decade fixed effects, ϕ_{st}, removes the presence of time-varying state-level supply and demand conditions that may jointly influence deep coal mine openings and relative coal capital. In Section 5.1, I show pre-trend tests for key covariates that support the inclusion of these controls. Because $\ln d_{it}$ is a county's contemporaneous distance to its nearest coal mine, π captures the contemporaneous effect of coal prices. Its inclusion removes any correlation between past and current coal prices, which supports a path dependence interpretation for lagged effects of $\ln d_i^0$.

Our coefficients of interest are β^τ. When $\tau > 0$, β^τ are lagged effects, with $\frac{\beta^\tau}{\pi} > 1$ and $\frac{\beta^\tau}{\pi} < 1$ implying strong and weak path dependence, respectively.[15] When $\tau < 0$, β^τ are lead effects and test for the presence of differential pre-trends in relative coal capital.[16] Equation (4) is my baseline specification. In Section 5.3, I also estimate variants of equation (4) to address remaining identification concerns.

4 Data

This section first details how spatial data on coal resources and mines are used to construct shocks to local coal transport distance. I then describe the construction of relative coal capital, my main outcome variable. For both variables, I present several tests to verify assumptions used in each construction procedure.

4.1 Local coal transport distance

The USGS National Coal Resource Assessment (NCRA) recently amassed and digitized spatial data on coal resources and mining that was previously held separately in the archives of state geological agencies (East, 2012). As shown in Figure A.3, the NCRA provides GIS

[14]Because mines serve multiple counties, county fixed effects also absorbs heterogeneity in the mill price set by the nearest mine right before the switch to deep coal. Heterogeneity in mill price captures variation in mines' marginal cost of coal extraction (Hotelling, 1929; C. d'Aspremont, 1979; Salop, 1979; Vogel, 2008), Hotelling rent associated with the size of its coal resource (Hotelling, 1931; Gaudet, Moreaux and Salant, 2001), and coal quality, such as heat, ash, and sulfur content, which can alter the amount of coal needed to produce a unit of electricity.

[15]In the absence of any path dependence, one may detect $\frac{\beta^\tau}{\pi} < 1$ if there was gradual capital depreciation over time.

[16]These lead effects provide indirect tests for whether the nearby availability of deep coal extraction was anticipated. Even if those shocks were predictable, it may not be the case that relevant forward-looking agents would act on it ex-ante. Miners cannot mine deep coal unless the technology to do so is available. Likewise, electric utilities are unlikely to construct coal-fired power plants in advance of nearby deep coal resources extraction.

shape files for coal resources of depths greater than and less than 200 ft from the surface for each of the major U.S. coal basins.[17] This paper makes several sample restrictions for data availability and estimation reasons.

I focus on coal resources found in the Illinois coal basin.[18] Of the major basins assessed by the NCRA, only for the Illinois Basin is there data on the location, area, and opening and closing years of coal mines.[19] This data goes back to 1890. The Illinois Basin is also advantageous for its geological properties. As Figure 2 shows, it has a dish-like shape with shallow coal resources in the outer regions and deeper resources near the center, implying that earlier shallow coal drilling in a given location does not geologically lead to deeper coal drilling, a potential source of endogeneity.[20] 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 of the five major coal basins, the Illinois Basin contains coal with the second lowest standard deviation in heat and ash content.

Within the Illinois Basin, I consider only large coal mines whose spatial area exceeds the 95% percentile for that basin. This is done to enable cleaner identification because large coal resources are more likely to be mined in response to regional, rather than local coal demand. Indeed FERC-423 forms reveal that counties with at least one of these large mines on average produce 3 times more coal than other counties in our sample region. Coal is also shipped further: the average quantity-weighted transport distance from these counties are 10 times longer. Additionally, larger coal resources have lower Hotelling, or scarcity rents, which may be an endogenous component of the delivered coal price (Hotelling, 1931). Figure 2 shows the location of these mines over the Illinois Basin. On the demand side, I restrict my county sample 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.

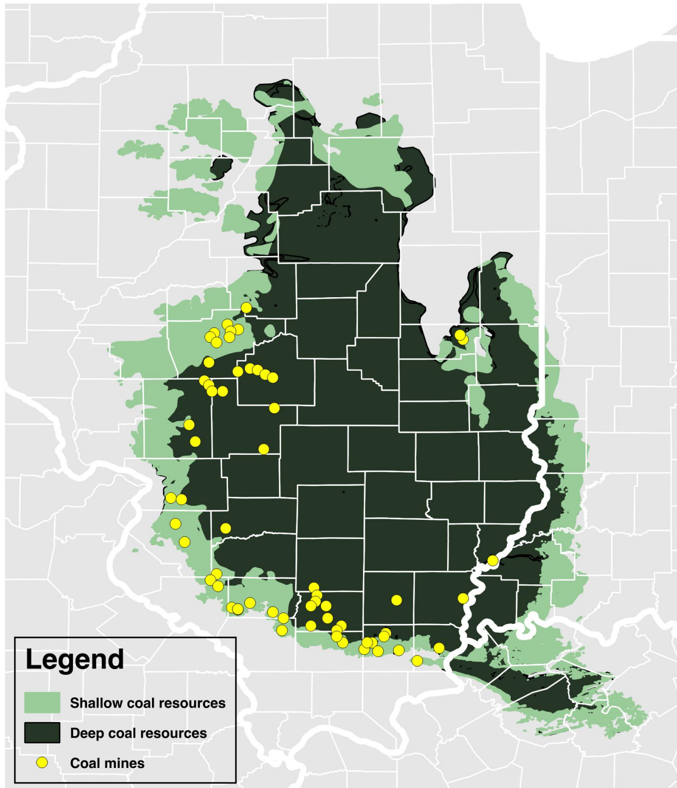
[17]The NCRA does not provide shape files with finer intervals of coal depth. Fortunately, Fisher (1910) notes 200 ft from the surface as generally the depth limit for manual coal mining prior to the 20th century.

[18]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.

[19]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.

[20]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 2 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 2 matches closely with that found in Campbell (1908).

Figure 2: 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.

The resulting sample of counties located in 11 states is shown in Figure A.4. These coal mine and county sample restrictions will be subjected to robustness checks. Figures 2 and A.4 visualize the regional scope of my coal supply shocks: less than 7% of sample counties have nearest sample coal mines in the same county.

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.5 maps county distance to nearest mine in 1890 and 1950, showing how its spatial distribution has changed during this period. 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.6 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.

I turn to three pieces of auxiliary evidence to verify my construction of coal transport distances. First, the use of transport distance assumes that transport costs are an important component of delivered coal prices. Figure A.7 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, 2024).

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.8 shows how the share of coal purchased by a county varies according to the rank in the transport distance to coal-supplying counties.[21] 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 local coal transport distance for sample counties. 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 thus provide confidence that my transport distance shocks is capturing variation in delivered coal prices. There is also no clear trend in these correlations.

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 for details). 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 for both 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

[21] Unfortunately, FERC-423 forms provide the county of origin for delivered coal, and not mine. Because I am unable to directly link power plants with coal mines, Figure A.8 provides a noisy test of the Herfindahl Principle.

appear in modern EIA-860 forms. To test the degree in which my constructed data may be missing retired power plants, I turn to historical data since 1920 which is only available at the national level (U.S. Census Bureau, 1975). Panel A of Figure A.9 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 the 1920-1970 period. Panel B of Figure A.9 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.9 draw 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.10 plots generating unit capacity reported in 2012 against the capacity reported in the late 1970s, showing a nearly one-for-one relationship. Table A.4 shows the distribution of reported primary fuel in 2012 conditional on primary fuel reported in the late 1970s. There has been little fuel switching since the late 1970s.[22]

The EIA-860 forms have one final limitation: they exclude generating units on power plants with less than 1 megawatt (MW) of combined electricity capacity. Appendix C discusses how this censoring may reduce the sample size, decreases the sample mean, and increases the skewness of relative coal capital. To address this, Appendix C details an imputation procedure which first predicts the number of missing power plants for each decade

[22] Note that there has been more fuel switching since 2010 (Energy Information Administration, 2020).

at the national level using a flexible polynomial function, estimated separately for coal and non-coal, and then distributes these predicted missing plants across electricity-producing counties. Robustness checks will examine various implementations of this procedure as well as its estimation uncertainty.

Even with imputed missing small power plants, the distribution of 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 depends on my imputation procedure. To produce estimates that are less sensitive to my imputation method, I estimate equation (4) using a Poisson model (Silva and Tenreyro, 2006).[23] I consider other models as robustness checks.

5 Reduced-form results

This section presents reduced-form evidence of path dependence in energy transitions. To guide the specification of my baseline model, I first test for pre-trends in key county covariates across models with different controls. I then present baseline reduced-form estimates and various robustness checks.

5.1 Examining pre-trends in county covariates

The baseline specification in equation (4) includes county and state-by-decade fixed effects. Table 1 verifies the importance of these controls by examining whether a county's distance to its nearest shallow mine before its switch to deep coal (i.e., $\ln d_i^0$) correlates with contemporaneous county characteristics that broadly capture coal demand. These covariates were obtained from Haines (2010) and shown down the rows of Table 1. Total and urban population proxy for residential coal demand. The number of establishments, employment,

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

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

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.

capital value, and output value for the manufacturing sector capture its coal demand.[24]

Table 1: Pre-trends in county covariates

Outcome	(1) Levels	(2) Changes	(3)
log population (1890-1990)	-0.13	-0.01	-0.02
	(0.09)	(0.01)	(0.01)
Counties	261	261	261
log urban population (1890-1980)	-0.09	-0.01	0.02
	(0.17)	(0.02)	(0.02)
Counties	183	171	171
log mgf establishments (1890-1990)	-0.20**	-0.05	0.05
	(0.10)	(0.06)	(0.06)
Counties	260	260	260
log mgf employment (1890-1990)	-0.16	0.11**	0.00
	(0.20)	(0.05)	(0.07)
Counties	258	255	255
log mgf capital stock (1890-1910)	0.07	0.17**	0.09
	(0.24)	(0.09)	(0.15)
Counties	106	105	105
log mgf output (1890-1940)	-0.44	0.17**	0.04
	(0.35)	(0.08)	(0.16)
Counties	114	113	113
State fixed effects	No	No	Yes

NOTES: Each coefficient comes from a separate regression of $\ln d_i^0$ on a contemporaneous county covariate as outcome. Time coverage for covariates indicated. County sample shown in Figure A.4. Covariates in column 1 in log levels. Covariates in columns 2 and 3 in first log differences. Column 3 augments column 2 by including state fixed effects. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

[24]A more precise test of pre-trends in manufacturing activity would isolate subsectors that are coal or electricity intensive. Historically, such data would have been collected by the U.S. Census of Manufacturers (COM). For our sample counties, pre-trend tests require data during 1880-1960. Unfortunately, as noted by Vickers and Ziebarth (2018), most COM data from this period has either since been lost or was never collected.

Each column of Table 1 estimates a model with different identifying assumptions. Column 1 shows the effect of $\ln d_i^0$ on log covariates in levels. Column 2 examines the effect of $\ln d_i^0$ on log covariates in first-differences, analogous to a panel estimator with county fixed effects. Column 3 augments column 2 by further including state fixed effects, akin to a panel estimator with county and state-by-decade fixed effects. Only in column 3 are pre-trends in these covariates systematically unrelated to $\ln d_i^0$. This confirms the importance of controlling for time-invariant county characteristics such as geography through county fixed effects, and time-varying state-level conditions through state-by-decade fixed effects, both of which are in the baseline statistical model shown in equation (4).

5.2 Baseline estimates of path dependence

The thick solid red line in Figure 3 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 2.

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 3 shows baseline 95% point confidence intervals for β^τ using county-level clustered standard errors. It is possible that error terms are spatially correlated across counties. To accommodate error correlation across a broader spatial scale, 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.[25] Figure A.13 augments the county and state-by-decade clustered standard errors in Figure 3 with estimated uncertainty from predicting missing small power plant by bootstrapping my imputation procedure (see Appendix C for details). Confidence intervals widen slightly.

I do not detect lead effects (i.e., $\beta^\tau : \tau < 0$). The absence of differential pre-trends in my outcome variable is consistent with the lack of pre-trends in county covariates shown in

[25] I further find very similar confidence intervals when using two-way clustered standard errors at the county and decade-levels to account for both serial correlation within a county and spatial correlation across counties in a decade. Those results are omitted because inference is complicated when there are only 11 decades in the sample.

Table 2: 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	-0.65	-0.34
	(0.67)	(2.32)
	—	—
1 decade lag	-0.67	-4.45**
	(1.25)	(1.90)
2 decades lag	-4.10***	-3.45
	(1.14)	(2.18)
3 decades lag	-3.74***	-7.74**
	(0.66)	(3.62)
4 decades lag	-3.50***	-5.66
	(0.71)	(5.24)
5 decades lag	-4.57***	-4.43
	(0.98)	(3.61)
6 decades lag	-3.70***	-2.87
	(0.75)	(3.90)
7 decades lag	-6.19***	-10.05**
	(1.37)	(4.21)
8 decades lag	-7.29***	-6.36
	(1.59)	(4.28)
9 decades lag	-7.29***	-3.78
	(1.57)	(4.57)
10 decades lag	-7.00***	-2.72
	(1.56)	(5.83)
$\ln d_{it}$ (π)	-1.53***	-3.47***
	(0.53)	(0.89)
Observations	2369	2369
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.4. 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.

column 3 of Table 1 and provides even stronger support for a parallel trends assumption.

Next, consistent with a reduced-form definition of strong path dependence, I detect statistically significant negative lagged effects that are larger in magnitude to the contemporaneous effect (i.e, $\frac{\beta^{\tau}}{\pi} > 1$). These negative lagged effects also intensify over time. Notably, Figure 3 shows stronger lagged effects two and seven decades later, which roughly coincides with the typical generator lifespan of four to six decades (see Appendix Figure A.12).[26] It is, however, still possible that the pattern shown in Figure 3 is due to differential retirement of existing coal and non-coal electricity capital. Column 2 of Table 2 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. Indeed, shocks to local coal transport distance alter relative coal capital investment over many later decades.

5.3 Robustness checks

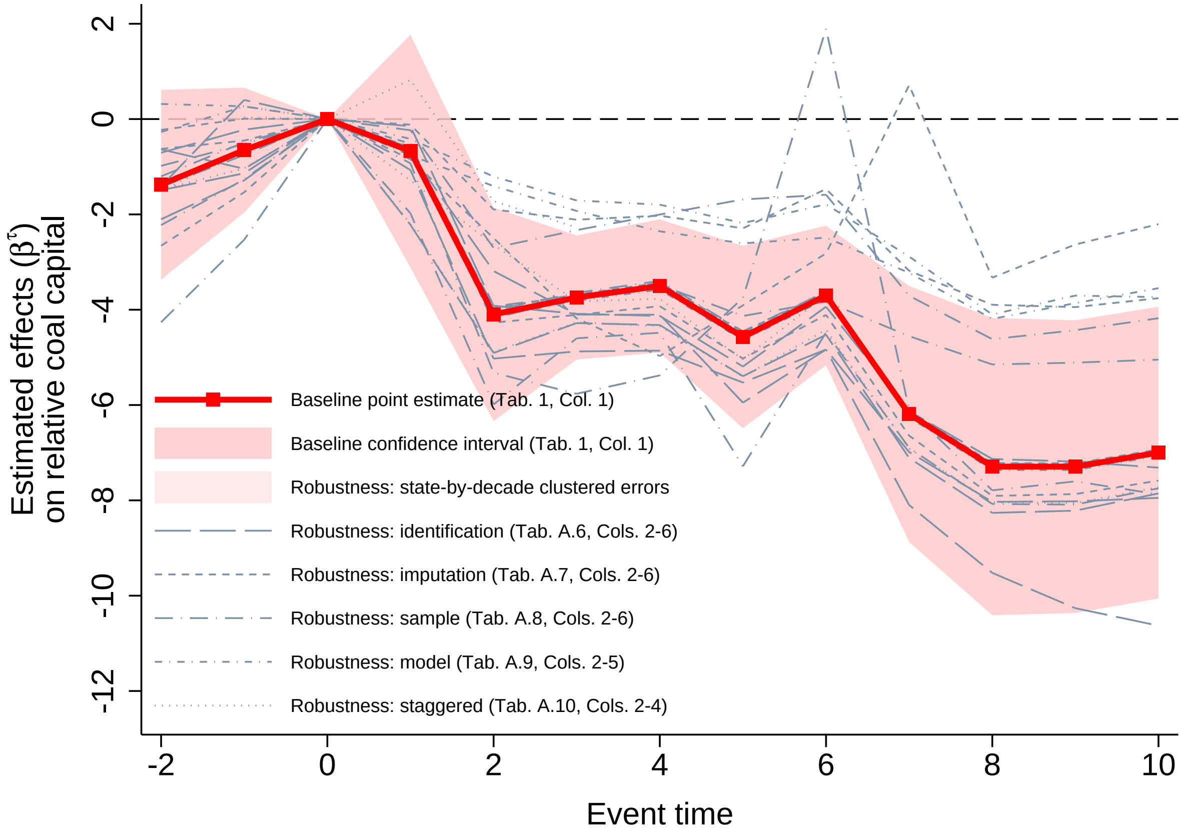
I turn now to a series of robustness checks designed to address further identification, data construction, sample restriction, and other statistical modeling concerns. Point estimates for all robustness checks are shown as thin non-solid blue lines in Figure 3.

Remaining identification concerns Table A.6 explores remaining identification concerns, with column 1 replicating my baseline estimates. It is possible that distance to the nearest shallow mine continues to directly influence relative coal capital after the switching event. While this implies that lagged effects would not have a clean path dependence interpretation, the resulting bias is likely towards zero, with lagged coefficients understating the true strength of path dependence. This is because, on its own, increasing mill prices from the ongoing depletion of the shallow mine would predict weakening, not strengthening, lagged effects over time.

One particular channel through which a shallow mine may exert direct influence on relative coal capital after the switching event is if the mine could price discriminate. For example, if mine A, the initial nearest supplier to a county, becomes the second nearest supplier 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

[26]Because generator retirement ages vary, the estimated lagged effects in Figure 3 do not pick up discrete new generator vintages but instead are averages across the distribution of generator lifespans.

Figure 3: 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.4. Time period is 1890-1990. Estimates also shown in column 2 of Table 2. 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 2-6 of Table A.6, columns 2-6 of Table A.7, columns 2-6 of Table A.8, columns 2-5 of Table A.9, and columns 2-4 of Table A.10.

a county's subsequent coal price even if it were no longer supplying coal.[27] To remove the possibility of price discrimination from affecting my results, column 2 estimates my baseline model without the counties for which the earlier supplying shallow mine becomes the second nearest mine at any point after the switching event. My results are little affected.

I next examine the possibility that distance to the shallow mine may be correlated with trends in other determinants of relative coal capital. Both the residential and manufacturing sectors consume electricity, while the manufacturing sector also consumes coal. In column 3,

[27] There is evidence from recent years showing price discriminating behavior by railroad companies that sell coal to power plants (Busse and Keohane, 2007; Preonas, 2024). Little is known, however, about whether such mark-ups existed across the longer time period considered in this study.

I augment the baseline model with three county-by-decade covariates available throughout the 20th century that proxy for the economic activity in these linked sectors: residential and manufacturing demand population, number of manufacturing establishments, and manufacturing employment, all in logs. Similarly, column 4 accounts for trends in transport costs by adding interaction terms between a linear time trend and two geographical features that affect transport costs: a county's distance to the nearest navigable waterway and its ruggedness, as approximated by its variance in slope, both in logs (see Appendix A for data details). Coefficients in columns 3 and 4 are similar to those of my baseline results. Despite these controls for local coal demand and transport costs, labor mobility across counties may still introduce endogeneity if remote coal mine openings draw away workers regionally affecting a given county's local wages and coal-fired electricity. Column 5 examines this possibility by controlling for county-by-decade log manufacturing wages, a county-level wage measure that is available for my sample period from Haines (2010), showing similar results. The current omitted event term marks the period before the initial switch to deep coal, which for some counties can span multiple decades. There may be differential pre-trends during this period that are not examined in my benchmark specification. To explore this possibility, column 6 redefines the focal event as just the single decade before the initial switch to deep coal such that pre-trends in the decades before this switch can be estimated. I do not detect such pre-trends.

Imputing missing small power plants Table A.7 examines alternative imputations for missing power plants with less than 1 MW capacity. Column 1 replicates my baseline result which uses a 4th order flexible polynomial function to fit national size distributions, separately for new coal and non-coal power plants built each decade, as detailed in Appendix C. Columns 2 and 3 show these results are stable to when 3rd and 5th order polynomial functions are used. If my imputation procedure is biased, one would want to weight my regression towards observations with more observed power plants. Column 4 does this by weighting each county-decade observation with each county's time-averaged number of power plants and shows similar results.[28] In column 5, I use a simpler, though less data-driven, imputation procedure by simply adding 1 MW capacity to coal and non-coal capital investment. The overall shape and precision of lagged effects are unchanged. While the magnitude of all coefficients are smaller, the ratio of lagged to contemporaneous effects, $\frac{\beta^\tau}{\pi}$, is consistently greater than one, indicative of strong path dependence. Column 6 estimates equation (4)

[28]These weights, however, may be endogenous and thus are not used in my benchmark specification.

using unadjusted relative coal capital. Despite the smaller sample size, lagged effects up to six decades have similar magnitudes to baseline results, with 3 to 5 decade lags statistically significant at the 10% level. Furthermore, consistent with strong path dependence, column 6 shows that lagged effects up to 9 decades later is larger in magnitude than the contemporaneous effect.

Column 7 provides an indirect test of my <1MW capacity imputation procedure by pretending 1-2MW capacity power plants are missing and instead impute these plants using the same procedure used to impute missing <1MW plants. To test my imputation procedure against observed data, column 7 estimates are compared to those in column 6, which includes observed 1-2MW plants and without imputation of <1MW plants. The lagged effects in columns 7 and 6 are of similar sign and magnitude for the first six decades. Longer lagged effects in column 7 are of smaller magnitude, though lagged effects remain larger than the contemporaneous effect up to 9 decades later, consistent with strong path dependence. This pattern, were it to hold for missing <1MW plants, suggests that my imputation of <1MW plants leads to lagged effects that are biased towards zero. Lastly, I detect a similar pattern when modeling unadjusted coal capital share, or $\frac{K_{cit}}{K_{cit}+K_{nit}}$, as shown column 6.[29]

Sample restrictions Table A.8 considers different sample restrictions, with column 1 reproducing my baseline results. On the coal supply side, recall that in order to isolate local coal transport distance shocks driven by regional conditions, I focused on large mines with areas above the 95th percentile of Illinois Basin mines. Estimates in columns 2 and 3 use transport distances constructed from mines with area above the 90th and 97.5th percentiles, respectively. While the inclusion of smaller mines result in statistically significant lead coefficients, these coefficients trend in the opposite direction of lagged effects.

On the coal demand side, to lessen the competing effects of other coal basins, my benchmark sample restricted counties to those within 250 miles of the nearest Illinois Basin coal resource and are situated closer to the Illinois Basin than to the Appalachian Basin. In column 4, I further weaken the influence of other coal basins by restricting the sample to counties within 200 miles of the nearest Illinois Basin coal resource. I find similar point estimates and standard errors. In column 5, I allow more counties into the sample by increasing the distance threshold to 300 miles. In column 6, I allow counties that are situated closer

[29] An alternative modeling approach may be to separately model the extensive margin of coal and non-coal capital investments via probit or logit models. In addition to having to address the missing small plant censuring issue, these models also do not account for the joint determination of coal and non-coal capital, which is essential for examining changes in fuel composition. Standard multinomial discrete choice models also do not readily account for simultaneous input choices.

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, point estimates are within the confidence intervals of baseline results.

Functional forms The log-log functional form of equation (4) implicitly assumes that the relationship between relative coal capital and coal transport distance is an isoelastic function. While this will be appealing for the theory in Section 6, this functional form may not be empirically supported. Figure A.14 examines whether the data supports linearity by estimating a variant of equation (4) that breaks log distance to the nearest mine into discrete bins, allowing a flexible relationship between distance and relative capital for each event-time period. Figure A.14 shows log relative coal capital predicted by log distance to the contemporaneous nearest mine, and by log distance to the earlier shallow mine two and seven decades after the switching event. These three relationships are approximately linear.

The baseline model in equation (4) includes 10 lag and 2 lead terms. More lag terms allow effects on new capital investments to be detected and 10 is the largest number of lags permitted by my 11 decade-long dataset. I vary the number of lead terms in Table A.9. Column 1 replicates the baseline model, column 2 includes 1 lead term, and column 3 includes 3 lead terms. I do not detect statistically significant lead terms across these models, nor very different lagged effects.

Finally, I consider two alternatives to the baseline 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 models overdispersion as a function of the expected outcome.

Column 4 of Table A.9 estimates equation (4) using a log-log linear model, while column 5 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, mirrors baseline results. While these coefficients are smaller in magnitude, the ratio of lagged to contemporaneous effects actually imply stronger path dependence.

Staggered treatment There are two concerns with dynamic staggered event study designs such as equation (4). First, the underlying composition of counties changes with event time as fewer sample counties are available for estimating longer lagged effects (Wing, Freedman and Hollingsworth, 2024; Baker et al., forthcoming). If county composition matters, one should see lagged effects changing significantly depending on the decade in which the switch occurs. To examine this, column 2 of Table A.10 estimates my benchmark specification on the subsample of counties that experience the switching event prior to the 1960s. Estimates are similar to the benchmark full sample estimates in column 1 of Table A.10. Column 3 implements a continuous version of this test by interacting the first five lagged terms – terms which can be estimated using early and late switching counties (see Figure A.6) – with the decade in which a county switches. These interaction terms are not statistically significant.

The second concern with staggered designs is the potential bias arising from using already-treated units as controls for later-treated units. To date, the econometrics literature has overwhelmingly focused on binary treatments in staggered designs (Callaway and Sant'Anna, 2021; Goodman-Bacon, 2021; de Chaisemartin and d'Haultfoeuille, 2020). My particular empirical approach – a staggered event study with dynamic effects, continuous treatment, and Poisson estimation – remains, to the best of my knowledge, beyond the scope of existing econometric results.[30] In the absence of established econometric methods, I informally address the improper controls concern by employing a stacked difference-in-difference approach following Wing, Freedman and Hollingsworth (2024). Specifically, I estimate a Poisson model after stacking the data into groups of counties by switching decade, limiting control counties to those that have not-yet or never experience the switching event, and further restricting my sample to counties with balanced event time observations (resulting in fewer lead and lag terms). Column 3 of A.10 shows qualitatively similar estimates to my benchmark results.

6 Mechanisms

Section 5's reduced-form evidence of strong path dependence in energy transition suggests it is possible for a permanent energy transition to be induced by a temporary intervention. This section combines theory and additional evidence to understanding the mechanism behind these results. To that end, I first develop a model of structural change model which incorporates two mechanisms that could operate at a county-level previously highlighted in the energy economics literature: increasing returns to scale and the accumulation of fuel-specific productivity. Informed by the structure of this model, I then conduct a series of

[30] Notable recent developments for staggered design with dynamic effects and continuous treatment in linear models are de Chaisemartin and d'Haultfoeuille (2024); Callaway, Goodman-Bacon and Sant'Anna (2024); de Chaisemartin and D'Haultfœuille (2025).

empirical tests designed to isolate the relevant mechanism. For completeness, I also test for mechanisms not considered by the model. While these tests by no means exhaust all possible explanations, their results lend confidence for the particular structural interpretation of my reduced-form results considered in Section 7.

6.1 A model of structural change for the electricity sector

A typical electricity producer operates several power plants, each consisting of fuel-specific generating units of different vintages. Generating units, in turn, combine one or more boilers and an electric generator to convert primary fuel, such as coal, into electricity. Each tier of this production structure - generating unit, power plant, producer - exhibits distinct features that have long been recognized in the energy economics literature. I first summarize these features before introducing them into a model.

F1 **"Putty-clay" capital** A generating unit consists of a boiler which turns fuel into steam and a generator which produces electricity from the boiler's steam. Boilers are typically designed to consume a particular fuel at a particular quantity. Sustained use of other fuels or use of the intended fuel in other quantities can lead to large efficiency losses (Avallone, Baumeister and Sadegh (2006), p.871). To account for this feature, prior literature has modeled electricity capital as "putty-clay," where it is fully utilized once built (Joskow, 1985, 1987; Atkeson and Kehoe, 1999).

F2 **Fuel and capital as perfect complements** Fuel is essential for producing electricity and cannot be substituted with other inputs. At the generating unit level, the combination of fuel and capital is typically modeled using a Leontief function (Komiya, 1962; Atkeson and Kehoe, 1999; Fabrizio, Rose and Wolfram, 2007).

F3 **Returns to scale** At the power plant-level, boilers can be linked to multiple generators. Thus, when a new boiler is installed, it can serve both new and existing generators, providing efficiency gains for generating units across multiple vintages of capital. This spillover effect, together with newer boilers typically being more efficient, provides a physical basis for increasing returns to scale at the power plant-level. Nerlove (1963) and Christensen and Greene (1976) provide seminal early estimates of scale economies in the electricity sector.

F4 **Imperfect substitutability across fuel-specific electricity** Electricity properties differ across input fuels. For example, electricity from coal provides "base load" supply that cannot be easily ramped up or down in response to variable demand, unlike electricity from natural gas. This imperfect substitutability across fuels is a crucial element of the directed technical change model developed by Acemoglu et al. (2012).

I incorporate these features into a model of electricity production with vintaged capital. In order to later map this model onto my reduced-form evidence, I consider a county that acts as a small open economy and served by a single electricity producer. The index t denotes the time increment between each capital vintage. At the top tier, the final good, electricity, Y_t, is generated using two intermediate goods, Y_{ct} and Y_{nt}, representing electricity from coal and all other fuels respectively. 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 long-run elasticity of substitution between electricity produced from the two intermediate sectors.[31] To allow for imperfect substitutability between electricity from different fuels (i.e., F4), I assume $\epsilon \in (1, +\infty)$. Final good price is normalized to 1.

Fuel-specific electricity, the intermediate good, is produced via middle and lower tiers of the production structure, corresponding to power plants and generating units respectively. 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 $\delta \in (0, 1)$ is the capital depreciation rate and $\alpha \in (0, 1)$ is the fuel-specific electricity elasticity of input. Returns to scale at the power plant-level (i.e., F3) is reflected in the middle tier where fuel-specific electricity comes from combining generating units across two vintages with scale parameter $\psi = 2\alpha$.[32] At the lowest tier, generating units combine fuel, E_{jt}, and t-vintaged capital, X_{jt} as perfect complements in a Leontief function (i.e., F2). $A_{Xjt} > 0$ and $A_{Ejt} > 0$ indicate 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 were the same across the two intermediate sectors in period $t-1$, such that $A_{Xct-1} = A_{Xnt-1}$ and $A_{Ect-1} = A_{Ent-1}$. Next, observe that when fuel and capital are perfect complements, efficient allocation in the lower production tier imply $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ for each intermediate sector j. Furthermore,

[31] This parameter is long-run in the sense that it reflects changes in capital.

[32] 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 3.

"clay" past-vintage capital (i.e., F1) with large efficiency losses implies that it is fully utilized, set at $A_{Xjt-1}(1-\delta)X_{jt-1} = A_{Ejt-1}E_{jt-1}$ for each intermediate sector j.[33] Rewriting the first order conditions in terms of current-vintage relative coal capital investment, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$, yields (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 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.[34] Equation (7) provides two channels through which past relative input prices, $\widetilde{w}_{t-1}$, affect current-vintage relative coal capital investment, $\widetilde{X}_t$. First, after applying equation (7) recursively, it can shown that past relative input prices affect past-vintage relative coal capital investment, $\widetilde{X}_{t-1}$. This occurs through the scale channel. 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 productivity channel would occur if there was accumulation over time in fuel-specific capital productivity such as via learning-by-doing, secular energy efficiency improvements, or if relative input prices direct research towards fuel-biased technological change.

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

6.2 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. Following Fabrizio, Rose and Wolfram (2007), I use power plant-level cost data from the Utility Data Institute (UDI) from 1981-1999.[35] To remove the influence

[33] Appendix D.2 shows how large efficiency losses when using electricity capital below its designed capacity lead to full utilization of past-vintage capital.

[34] I choose this notation to be consistent with that found in Acemoglu et al. (2012).

[35] Ideally, to be consistent with data on relative coal capital, plant-level cost data should also cover the 20th century. Unfortunately, cost data prior to 1981 is not readily available. Use of 2009-2012 annual averages to construct productivity measures potentially assuages a concern that locations with higher coal generation share may also have higher coal generation volatility due to less available dispatchable energy sources (e.g., oil and gas).

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.3 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 3 displays estimates of ψ for coal-only power plants in my baseline county sample. I estimate a scale parameter of $\hat{\psi} = 1.8$ 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, shocks to local coal transport 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 transport distance 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 shocks, or the slope of the lagged effects shown in Figure 3 over time. Column 2 of Table 3 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), assuaging concerns over a seemingly low Kleibergen-Paap first-stage F-statistic. Henceforth, my preferred scale parameter estimate is $\hat{\psi} = 1.66$ from column 2.[36]

Table 3: 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***
	(0.03)	(0.09)	(0.05)	(0.05)
Implied scale parameter, ψ	1.78***	1.66***	1.75***	1.96***
	(0.09)	(0.24)	(0.16)	(0.21)
Kleibergen-Paap F-stat		3.25		
CLR p-value		0.01		
CLR confidence int. (90%)		[0.54, 0.89]		
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
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.4. 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 shows the Kleibergen-Paap F-statistic (Kleibergen, 2004) as well as the p-value and confidence interval (in brackets) from a conditional likelihood ratio test against a null hypothesis that $1/\psi$ is zero (Moreira, 2003). Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

The production function in Section 6.1 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 3. 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 the entire U.S. and to plants UDI designates as consuming natural gas and oil as primary fuels.[37] I find scale parameters from

[36]By comparison, 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.

[37]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

gas and oil-fired power plants that are statistically indistinguishable from that of coal-fired power plants.

6.3 Testing for productivity effects at the generating unit level

Past shocks to local coal transport distance could also directly affect the accumulation of coal-specific productivity. If so, the IV estimator in Section 6.2 would not satisfy the exclusion restriction needed for identifying the scale parameter. To test whether productivity effects violate the exclusion restriction assumption, I turn to generating unit-level regressions where productivity effects can be more cleanly isolated. I employ two standard measures of generating-unit productivity. Following Davis and Wolfram (2012), 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,[38] I estimate the following regression for generating unit g, in power plant p, located in county i and state s with both productivity measures as outcomes

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

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 transport distance shocks on generating unit productivity.

Results are shown in Table 4 for generating units in coal-only power plants located in my baseline county sample. Column 1 examine log capital productivity while column 2 examine log thermal efficiency. For neither outcome do I detect statistically precise effects. However, the uncertainty in these estimates also cannot rule out economically meaningful productivity effects.[39]

fuel consumed and then assigning primary fuel to the fuel with the highest energy input.

[38]Ideally, to be consistent with data on relative coal capital, generating unit-level productivity variables should also cover the 20th century. Unfortunately, construction of these productivity variables require boiler-to-generator correspondences which were only made available in 2009. See Appendix A for details. Because productivity presumably accumulates over time, any lagged effects of past shocks to local coal transport distance should be stronger and more detectable with more recent productivity data.

[39]Because past coal prices affect relative productivities of coal and non-coal generators in the presence

Table 4: Productivity regressions at the generating unit-level

	(1) ln capital prod.	(2) ln fuel prod.
$\ln d_i^o \times sinceEvent_i$ (ω_1)	-0.052	0.015
	(0.032)	(0.047)
$\ln d_i^o$ (ω_2)	0.12	-0.33
	(0.17)	(0.26)
$sinceEvent_i$ (ω_3)	0.14	-0.12
	(0.12)	(0.18)
Generating units	235	235

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.4. Outcome in column 1 is the 2009-2012 average ratio of electricity generation to generating unit capital, which approximates capital productivity, A_{Xct}. Outcome in column 2 is the 2009-2012 average thermal efficiency, which approximates fuel productivity, A_{Ect}. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

It is important to note that failing to find local productivity effects from local coal transport distance shocks does not necessarily imply there are no productivity effects at a more aggregate level. Induced innovation can spillover across the research sector and is not confined within counties. Similarly, if productivity improvements occur through human capital accumulation, there could be spillovers across counties through labor reallocation by firms with multiple power plants. I return to this issue in Section 7.

6.4 Alternative mechanisms

The model in Section 6.1 necessarily omits other potential mechanisms. While the presence of other mechanisms would not invalidate Section 5's reduced-form estimates of strong path dependence in energy transitions, they would complicate how these results are structurally interpreted.

of directed technical change, a more comprehensive test would be to jointly evaluate effects on coal and non-coal generator productivities. Unfortunately, our productivity measure is only suitable for fossil fuel generators and there are a limited number of non-coal (i.e., oil and natural gas) generators in Table 4's estimating sample.

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.[40] 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.11 which estimates my baseline model in equation (4) but only keeps the subset of observations before there was state PUC regulation of electric utilities (see Appendix A for details). Despite the smaller sample size, I detect statistically significant lagged effects with magnitudes that are similar to my baseline estimates in column 1 of Table A.11.

A second important regulation is the U.S. Clean Air Act. Beginning with the 1970 U.S. Clean Air Act Amendments, 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 were 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.11 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. I find lagged effects that are similar in magnitude to my baseline results. Column 3 also suggests that my path dependence estimates do not differ 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.

[40] 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.

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 strong 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 (2022) 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.[41] To examine whether past coal transport distance shocks affect the coal transport sector itself, I turn to a county-level version of the specification in equation (10). Column 1 of Table A.12 examines whether there are event time-varying effects of past transport distance 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.12 also fails to detect effects of past transport distance shocks on log highway density in 2010 (in miles per square mile).

I also consider downstream effects. Previous literature detects long-term effects of historical access to electricity on subsequent local manufacturing sector employment (Kline and Moretti, 2014), population density (Severnini, 2023), and energy efficiency (Hawkins-Pierot and Wagner, 2025). 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 other fuels. If so, one would expect my lagged effects to be altered when I control for local manufacturing sector demand. Column 3 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

[41] 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).

into historically coal-dependent locations, these residents may support more lenient local environmental policies that enable the expansion of coal-fired electricity even as economic circumstances that initially favored coal disappear. Using the same specification from columns 1 and 2, columns 3 and 4 of Table A.12 explore this sorting mechanism by estimating event time-varying effects of past coal transport distance shocks on the share of a county's population belonging to one of three major environmental NGOs in 1996 and the county share of eligible voters that voted for the 2000 Republican Presidential candidate (see Appendix A for data details). Neither proxy for environmental preferences responds to past transport distance shocks.

Altogether, none of the mechanisms examined in Table A.12 yield statistically precise effects. But as with the productivity estimates in Table 4, one can neither rule out potentially meaningful effects given uncertainty in these estimates, an important caveat for the structural interpretation and simulations presented in Section 7.

Capital adjustment costs Finally, path dependence in relative coal capital can arise from convex capital adjustment costs making it more costly to adjust higher levels of capital stock. For example, if a county has high baseline relative coal capital stock, a positive, temporary shock to relative coal prices would exhibit a slower transition away from coal in the presence of convex capital adjustment costs.

Appendix D.4 augments my baseline model introducing convex adjustment costs in the level of new-vintage capital yielding three empirical implications. First, because the producer's first order condition (i.e., equation (A.17)) no longer delivers a closed-form log-linear relationship between current relative coal capital and current relative input prices, it is harder to estimate in a linear regression framework. Second, because convex adjustment costs and the elasticity of substitution both influence how relative capital responds over time to changes in relative input prices, it becomes difficult for an empirical framework to separately identify these parameters without committing to particular nonlinear functional forms and identification off those functional form assumptions. Lastly, in the presence of convex adjustment costs, any recovery of the elasticity of substitution based on reduced-form regressions of relative coal capital investment on relative input prices that ignore these frictions will generally understate the true elasticity of substitution: reduced-form estimates combine the effect of the underlying substitution parameter with the transition-dampening effect of the convex capital adjustment costs. Thus, presence of such adjustment costs, while hard to empirically validate, adds another caveat to the structural interpretation and simulations in Section 7.

7 Structural interpretation

This section employs the structure of the model in Section 6.1 to interpret my reduced-form estimates. My model focuses on returns to scale as the operating mechanism given the statistical precision of the scale parameter estimate in Table 3. I view this as an exploratory exercise, as it implicitly, and perhaps incorrectly, assumes that other mechanisms, such as the ones that are noisily estimated in Sections 6.3 and 6.4 and capital adjustment costs, do not play a role.

First, I formally define scale-driven path dependence in energy transitions as a function of two parameters: returns to scale, ψ, and the long-run elasticity of substitution between electricity produced from coal and other fuels, ϵ. I then use this definition to recover ϵ implied by my reduced-form estimates of path dependence in Section 5 and my estimates of the scale parameter in Section 6.2. Finally, I explore through calibrated model simulations the conditions under which temporary interventions to fuel composition can induce sustained future transitions away from coal at the county-level. Further simulations explore implications for U.S.-wide energy transitions away from coal.

7.1 Formal definitions of scale-driven path dependence

I formally defined scale-driven path dependence within the model in Section 6.1. To do this, I first apply a log transformation to equation (7) and rewrite it recursively. Current-vintage relative coal capital investment becomes

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

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 scale-driven path dependence in energy transitions. Formally,

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

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

Strong scale-driven 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, each of which I first consider in isolation.

The first force is increasing returns to scale, captured by ψ. To isolate this mechanism, suppose there is only one intermediate sector that uses coal to produce electricity. When $\psi = 2\alpha > 1$, there is a "push" from cross-vintage scale effects: an additional unit of past coal capital boosts the marginal product of current coal capital more than an additional unit of current coal capital lowers own marginal product from within-vintage diminishing returns due to $\alpha \in (0, 1)$. This enables relative coal prices to have a stronger effect on future coal capital investment than on current coal capital investment.[42]

However, when there is more than one sector, $\psi > 1$ alone does not dictate the strength of scale-driven 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 scale-driven path dependence can only be achieved when increasing returns to scale provides a large enough push to overcome the pull from imperfect substitutability, or when $\psi > \frac{-\epsilon}{1-\epsilon}$.[43] In the context of

[42] 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, $\frac{\partial^2 Y_{ct}}{\partial X_{ct} \partial X_{ct-1}} > -\frac{\partial^2 Y_{ct}}{\partial^2 X_{ct}^2}$ when $\psi > 1$. Thus, when $\psi > 1$, cross-vintage increasing returns to scale in the cross derivative dominate within-vintage diminishing returns in the own second derivative with respect to current capital.

[43] "Push" and 'pull" mechanisms can be found in models of directed technical change (Acemoglu, 2002; Acemoglu et al., 2012).

the electricity sector, 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, such as electricity reliability.

7.2 Recovering the long-run elasticity of substitution

I recover ϵ using the mapping between the structural expression in equation (11) and reduced-form coefficients in equations (4) and (8).[44]

To start, observe that 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. To convert from event-time to vintage-time, I weight each reduced-form lagged effect, $\hat{\beta}^\tau$, by the probability of generator retirement, ι^τ, from the age distribution of retired generators in my sample. Specifically, using $\hat{\beta}^\tau$ and $\hat{\pi}$ from column 2 of Table 2, the ratio of reduced-form lagged to contemporaneous effect is $\widehat{\rho} = \frac{\sum_{\tau=1}^{5} \hat{\beta}^\tau * \iota^\tau}{\hat{\pi}} = 1.45$, with a standard error of 0.72.[45] Mapping this ratio to its structural analog $\frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_{t-1})} / \frac{\partial \ln(\widetilde{X}_t)}{\partial \ln(\widetilde{w}_t)} = \frac{\alpha(1-\epsilon)\epsilon}{(\varphi-1)^2} / \frac{\epsilon}{(\varphi-1)}$ and taking the estimated scale parameter $\hat{\psi} = 1.66$ from column 2 of Table 3, one can express the long-run elasticity of substitution entirely in terms of estimated statistical parameters

$$\epsilon = 1 + \frac{\widehat{\rho}}{\frac{\hat{\psi}}{2} - \widehat{\rho}(1 - \frac{\hat{\psi}}{2})} = 3.5$$

7.3 Simulating future energy transitions away from coal

Strong scale-driven path dependence implies it is possible for a temporary shock to fuel composition to induce permanent fuel switching. In the context of climate policy, it suggests that a temporary policy can induce a long-term decline in carbon emissions at the county-level. But under what conditions? In particular, given the abundance of coal resources in the U.S., what is the required magnitude and/or duration of an intervention that can trigger a sustained energy transition away from coal?

To shed light on these questions, I calibrate my structural model with reduced-form parameters to simulate future county-level electricity sector CO_2 emissions following rela-

---

[44] The long-run elasticity of substitution between fuel inputs, ϵ, appears across a recent class of optimal climate policy models (Acemoglu et al., 2012; Golosov et al., 2014; Lemoine, 2024; Fried, 2018; Acemoglu et al., 2023). However, substitution elasticities in those papers differ conceptually from that of this paper (i.e., coal-non coal vs. dirty-clean energy substitution; county vs. national vs. global aggregation), limiting any direct comparison.

[45] Since $\iota^\tau \in \{.38, .27, .22, .11, .02\}$, $\widehat{\rho} = \frac{\sum_{\tau=1}^{5} \hat{\beta}^\tau * \iota^\tau}{\hat{\pi}} = \frac{-(4.47*.38+3.46*.27+7.75*.22+5.68*.11+4.46*.02)}{-3.48} = 1.45$.

tive coal price shocks of varying magnitude and duration. To ground my simulations in recent economic conditions, the magnitude of the shocks considered are benchmarked to recent average national relative coal prices following the introduction of natural gas hydraulic fracturing. Figure A.15 shows that as a consequence of hydraulic fracturing, relative U.S. coal prices in 2009 and 2010 was 143% higher than what a quadratic trend estimated over 1985-2008 would have predicted.

My simulations employ several simplifying assumptions (see Appendix E for details). First, only electricity from coal and natural gas are considered.[46] Second, to avoid forecasting trends in relative productivity, electricity demand, and other economic conditions, I assume that future relative productivity, total coal and natural gas capital, capital depreciation, and baseline relative coal prices in the absence of the shock are held constant at recent average county values. Third, I assume that the scale and elasticity of substitution parameters are constant throughout the simulation period. Because of these assumptions, one should not interpret these simulations as forecasts of county-level electricity sector CO_2 emissions but rather as exercises in understanding the conditions for triggering a sustained future energy transition away from coal.

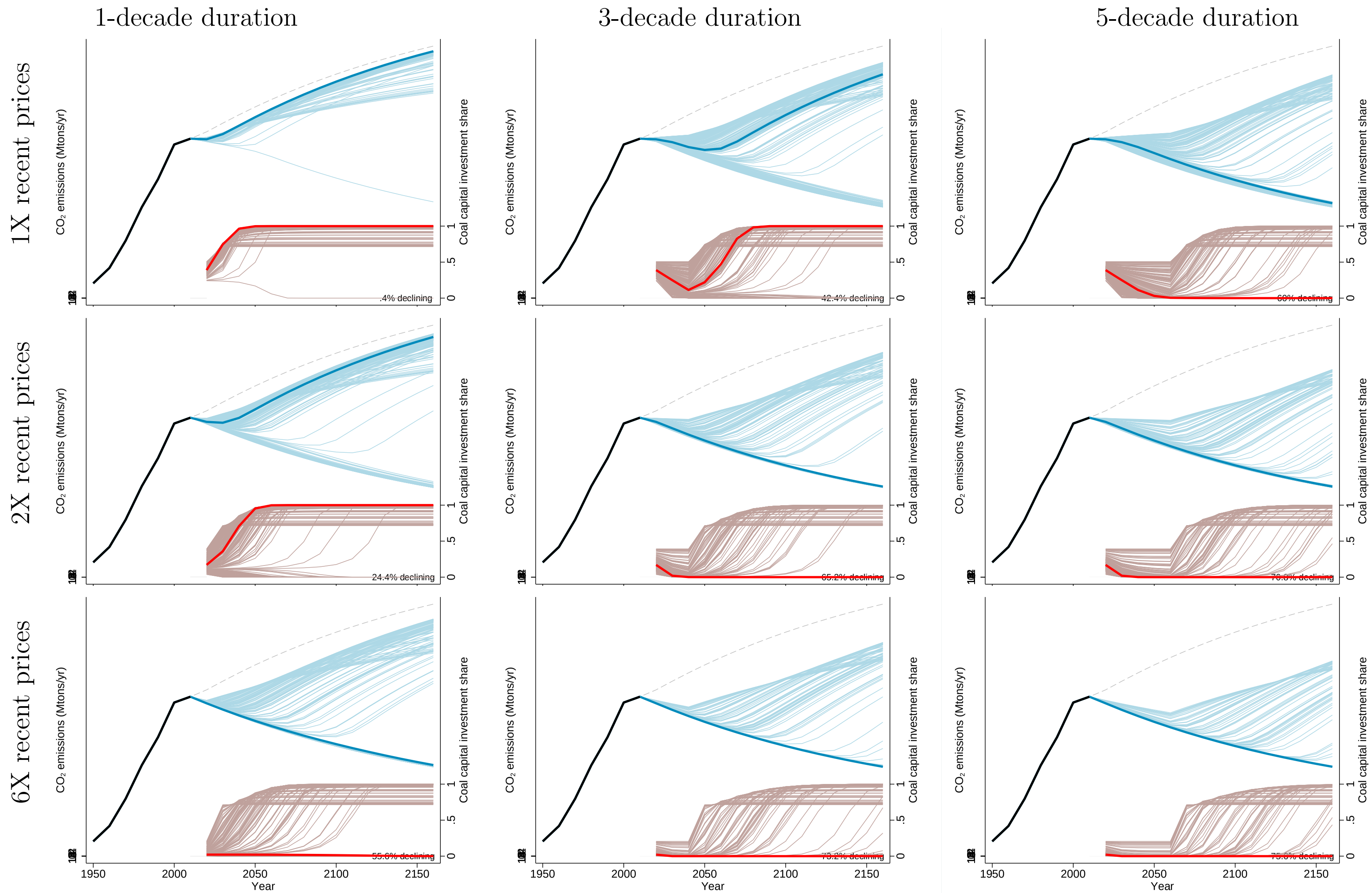
Figure 4 shows how CO_2 emission trajectories are altered when one varies the magnitude and duration of the temporary relative coal price shock. Top, middle, and bottom panels use price shocks that are 1, 2, and 6 times that of recent relative coal prices. Left, middle, and right panels use price shocks that last 1, 3, and 5 decades. Business-as-usual emissions in the absence of the shock are 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 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 uncertainty for each structural parameter. Each panel also indicates the percentage of draws in which CO_2 emissions are weakly declining in the long-term.[47]

In general, the likelihood of achieving weakly declining CO_2 emissions in the long-term

[46] Initializing the simulations requires current relative coal to non-coal input prices. Input prices for the other major U.S. electricity energy sources are hard to obtain, i.e., water for hydro dams, uranium/plutonium for nuclear plants, marginal land for solar and wind, and oil provides less than 1% of recent U.S. electricity generation. By projecting only the coal to natural transition, I am likely under counting non-coal electricity capacity investment and thus overstating the coal capital investment shares in Figure 4. However, projected total electricity CO_2 emissions are likely unaffected as other major non-coal energy sources do not emit CO_2.

[47] Because natural gas still contains carbon, in none of the simulations do CO_2 emissions reach zero in the long term. Instead, emissions converge asymptotically to a steady-state level where all electricity is produced from natural gas. Thus, these simulations focus on whether temporary policies achieve weakly declining CO_2 emissions in the long-term.

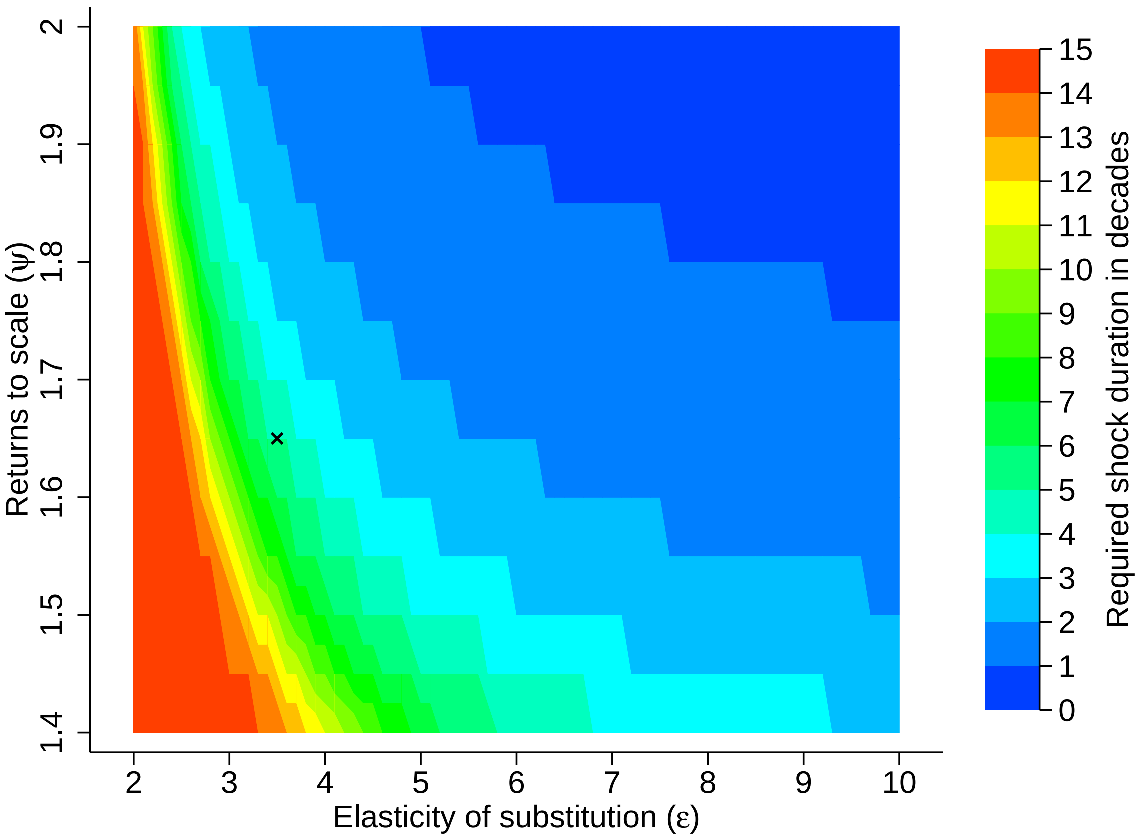
Figure 4: Simulating future county-level CO_2 emissions following temporary relative coal price shocks



NOTES: Simulations of CO_2 electricity sector emissions over 2010-2150 for the average sample U.S. county following temporary relative coal price shocks of varying magnitudes and durations. Top, middle, and bottom panels use price shocks 1, 2, and 6 times that of recent relative coal prices (see Figure A.15). Left, middle, and right panels use price shocks that last 1, 3, and 5 decades (shown as shaded gray areas). Solid black lines show historical emissions during 1950-2000. 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 Mton/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 uncertainty for ψ and ϵ. Percentage of draws with weakly declining long-term CO_2 emissions are also shown. See Appendix E for details.

increases with larger shocks and/or shocks of longer duration. Examining the top row of Figure 4, if recent relative coal prices were to last up to 3 decades, CO_2 would still rise in the long-term on average. 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 5 decades. Going down the first column of Figure 4, if an intervention can only last for 1 decade, it must be at least 6 times that of recent relative coal prices for a greater than 50% chance of achieving weakly declining long-term emissions.

Figure 5: 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 a shock equal in magnitude to that of 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.

Figure 4 suggests that given historical parameter values, a sustained energy transition at the county level would require either a very large and/or long duration intervention. How would different parameter values alter these requirements? This is of interest for two reasons. First, future electricity production may exhibit different returns to scale as thermal limits are reached. Second, expansion of electricity grids may alter the degree of substitutability across fuels.[48]

Figure 5 explores the energy transition implications of varying push and pull forces within the structure of my model by considering different values of ϵ and ψ. For each pair of parameter values, the heat map in Figure 5 shows the minimum number of decades a shock equal in magnitude to that of 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 supply driven by the changing regional accessibility of subsurface coal, I find that a negative shock triggers a declining trajectory in the relative use of coal for electricity lasting for up to ten decades. This suggests that it is possible for temporary shocks of sufficient magnitude and/or duration to induce permanent switch away from coal in the electricity sector. Additional evidence suggests increasing returns to scale in electricity production as the underlying mechanism. To interpret these results, I develop a model of scale-driven structural change for the electricity sector which allows for a formal definition of strong path dependence and a mapping between my reduced-form estimates and a key structural parameter, the long-run elasticity of substitution between coal and other fuels.

This historical evidence is particularly timely given recent developments in the U.S. electricity sector and increasing concerns over climate change impacts. The recent 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 calibrated structural model demonstrate that a sustained energy transition away from coal would require either a larger and/or longer lasting shock to fuel composition. These simulations also show that a sustained energy transition away from coal can be more easily achieved if investments were made to either increase scale economies in electricity production or towards stronger substitutability

[48]While renewable energy is not considered in my simulations, storage technology can improve the reliability of solar and wind-based electricity, making renewable energy sources stronger substitutes for coal.

between electricity from coal and other fuels.

The presence of strong path dependence in energy transitions also provides further support for externality pricing over second-best policies that favor specific technologies or resources. It is widely argued that technology- and resource-specific policies are inefficient because they may fail to target cost-effective mitigation strategies. The presence of path dependence amplifies this cost. Suppose a natural gas-specific subsidy induces a transition from coal to natural gas but large climate damages ultimately requires a switch to even cleaner fuels. The ensuing path dependence in natural gas would make the eventual switch to cleaner fuels more costly. A carbon price would avoid this detour.

References

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

Acemoglu, Daron, and Will Rafey. 2023. "Mirage on the horizon: Geoengineering and carbon taxation without commitment." *Journal of Public Economics*, 219: 104802.

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, Philippe Aghion, Lint Barrage, and David Hémous. 2023. "Climate change, directed innovation, and energy transition: The long-run consequences of the shale gas revolution." National Bureau of Economic Research.

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.

Aghion, Philippe, Lint Barrage, Eric Donald, David Hémous, and Ernest Liu. 2025. "Transition to Green Technology along the Supply Chain." National Bureau of Economic Research Working Paper 33934.

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): 313–338.

Atack, Jeremy. 2015. "Historical Geographic Information Systems (GIS) database of Steamboat-Navigated Rivers During the Nineteenth Century in the United States."

Atack, Jeremy. 2017. "Historical Geographic Information Systems (GIS) database of Nineteenth Century U.S. Canals."

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

Atkinson, Scott E, and Robert Halvorsen. 1976. "Interfuel substitution in steam electric power generation." *Journal of Political Economy*, 84(5): 959–978.

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.

Baker, Andrew, Brantly Callaway, Scott Cunningham, Andrew Goodman-Bacon, and Pedro HC Sant'Anna. forthcoming. "Difference-in-differences designs: A practitioner's guide." *Journal of Economic Literature*.

Bank, World. 2014. "World Development Indicators." The World Bank.

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.

Beach, Brian, and W Walker Hanlon. 2018. "Coal smoke and mortality in an early industrial economy." *The Economic Journal*, 128(615): 2652–2675.

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.

Callaway, Brantly, and Pedro HC Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C Sant'Anna. 2024. "Difference-in-differences with a Continuous Treatment." , (32117).

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

Casey, Gregory. 2024. "Energy efficiency and directed technical change: implications for climate change mitigation." *Review of Economic Studies*, 91(1): 192–228.

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. 2024. “Canary in a coal mine: Infant mortality and tradeoffs associated with mid-20th century air pollution.” *Review of Economics and Statistics*, 106(3): 698–711.

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.

de Chaisemartin, Clément, and Xavier d’Haultfoeuille. 2024. “Difference-in-differences estimators of intertemporal treatment effects.” *Review of Economics and Statistics*, 1–45.

de Chaisemartin, Clément, and Xavier d’Haultfoeuille. 2020. “Two-way fixed effects estimators with heterogeneous treatment effects.” *American Economic Review*, 110(9): 2964–2996.

de Chaisemartin, Clément, and Xavier D’Haultfœuille. 2025. “Treatment-Effect Estimation in Complex Designs under a Parallel-trends Assumption.”

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.” U.S. Energy Information Administration.

Energy Information Administration. 2012. “Annual Energy Review, 1982-2011.” U.S. Energy Information Administration.

Energy Information Administration. 2020. “Electric Power Annual.” U.S. Energy Information Administration.

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.

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

Fouquet, Roger. 2016. “Path dependence in energy systems and economic development.” *Nature Energy*, 1(8): 16098.

Fried, Stephie. 2018. “Climate Policy and Innovation: A Quantitative Macroeconomic Analysis.” *American Economic Journal: Macroeconomics*, 10(1): 90–118.

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, GŽrard, 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.

Golosov, Mikhail, John Hassler, Per Krusell, and Aleh Tsyvinski. 2014. “Optimal Taxes on Fossil

Fuel in General Equilibrium." *Econometrica*, 82(1): 41–88.

Goodman-Bacon, Andrew. 2021. "Difference-in-differences with variation in treatment timing." *Journal of econometrics*, 225(2): 254–277.

Griffin, James M, and Paul R Gregory. 1976. "An intercountry translog model of energy substitution responses." *The American Economic Review*, 66(5): 845–857.

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. 2020. "Coal smoke, city growth, and the costs of the industrial revolution." *The Economic Journal*, 130(626): 462–488.

Hassler, John, Per Krusell, and Conny Olovsson. 2021. "Directed technical change as a response to natural resource scarcity." *Journal of Political Economy*, 129(11): 3039–3072.

Hawkins-Pierot, Jonathan T, and Katherine RH Wagner. 2025. "Technology lock-in and costs of delayed climate policy." *Journal of Political Economy Microeconomics, forthcoming.*

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.

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. 2022. "Regulatory Induced Risk Aversion in Coal Contracting at US Power Plants: Implications for Environmental Policy." *Journal of the Association of Environmental and Resource Economists*, 9(1): 51–78.

Jo, Ara. 2025. "Substitution between clean and dirty energy with biased technical change." *International Economic Review*, 66(2): 883–902.

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 *Handbook of Industrial Organization.* Vol. 2, 1449 – 1506. Elsevier.

Kleibergen, Frank. 2004. "Testing Subsets of Structural Parameters in the Instrumental Variables." *The Review of Economics and Statistics*, 86(1): 418–423.

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. 2019. "Natural Gas Prices and Coal Displacement: Evidence from Electricity Markets." *International Journal of Industrial Organization.*

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.

Lanzi, Elisa, and Ian Sue Wing. 2010. "Directed Technical Change in the Energy Sector: an Empirical Test of Induced Directed Innovation." *mimeo.*

Lemoine, Derek. 2024. "Innovation-led transitions in energy supply." *American Economic Journal: Macroeconomics*, 16(1): 29–65.

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.

Papageorgiou, Chris, Marianne Saam, and Patrick Schulte. 2017. "Substitution between clean and dirty energy inputs: A macroeconomic perspective." *Review of Economics and Statistics*, 99(2): 281–290.

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

Pindyck, Robert S. 1979. "Interfuel substitution and the industrial demand for energy: an international comparison." *The Review of Economics and Statistics*, 169–179.

Preonas, Louis. 2024. "Market power in coal shipping and implications for US climate policy." *Review of Economic Studies*, 91(4): 2508–2537.

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

Severnini, Edson. 2023. "The power of hydroelectric dams: historical evidence from the United States over the twentieth century." *The Economic Journal*, 133(649): 420–459.

Silva, J. M. C. Santos, and Silvana Tenreyro. 2006. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

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

Stern, David I. 2012. "Interfuel Substitution: A Meta-Analysis." *Journal of Economic Surveys*, 26(2): 307–331.

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.

Vickers, Chris, and Nicolas L Ziebarth. 2018. "The Census of Manufactures: An Overview." *Handbook of Cliometrics*, 1–24.

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.

Wing, Coady, Seth M Freedman, and Alex Hollingsworth. 2024. "Stacked Difference-in-Differences." National Bureau of Economic Research Working Paper 32054.

Appendix to

Estimating Path Dependence in Energy Transitions

FOR ONLINE PUBLICATION

A Data Sources

This section details data used in the paper.

A.1 Coal resources, mining, and delivered prices

USGS National Coal Resource Assessment (NCRA) I use two spatial datasets from the NCRA (East, 2012). The first dataset contains 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 Figure A.4 which maps coal resources for the two basins by depth. The second dataset contains characteristics of all coal mines in the Illinois Basin that has operated since 1890. Variables include mine location, opening year, closing year, and area.

Construction of local coal transport distances requires several 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 2. 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.[49] 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, noting whether it is from a shallow or deep coal resource. This distance is d_{it} in equation (4). Finally, for each county, I find the first decade in which a county's nearest mine first switches from shallow coal to deep coal. Distance to the nearest shallow mine right before the switching event is denoted d_i^0 in equation (4).

FERC-423 forms FERC-423 provides annual data on the quantity, price, heat content, sulfur content, and ash content of purchased coal for each pair of purchasing power plant and county of coal origin. This paper uses FERC-423 data in four ways. First, for the 1990-1999 period when the county of coal origin is more reliable,[50] I calculate county-level average annual coal production and quantity-weighted export distance. I then compare average coal production and export distance between counties with large mines and those without within my sample region. Second, I use FERC-423 data on coal heat, sulfur, and heat content, 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. Third, data from the 1990-1999 FERC-423 forms are used to test the Herfindahl Principle, as shown in Figure A.8. Fourth, I use observed delivered coal prices from the entire set of 1972-1999 FERC-423 forms to verify my local coal transport distances in Table A.2.

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

[50]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.[51] 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 1.

Underlying assumptions behind this data construction procedure are examined in Figures A.9, A.10 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.

Knowing when generating units were built, I can also construct a county-by-decade dataset of relative coal capital investment. 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 Control variables

Residential and manufacturing sector covariates County-by-decade residential and manufacturing covariates from 1890 to 1990 come from historical U.S. censuses, collected by Haines (2010).[52] These variables include total population, urban population, number of manufacturing establishments, manufacturing employment, manufacturing capital value, and manufacturing output value. 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.)[53] and is a modification of the method used by Hornbeck (2012). The resulting data serve as outcome variables in the pre-trend tests shown in Table 1 and as control variables in column 3 of Table A.6.

[51] 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.

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

[53] Available: https://www.nhgis.org{}/documentation/gis-data

Geographical covariates County-level variance in slope is constructed using the USGS National Elevation Dataset. County-level distance to nearest navigable river or canal at the start of the 20th century combines GIS shape files from Atack (2015) and Atack (2017).[54] Both variables are used as controls in column 4 of Table A.6.

A.4 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 3.

EIA-923 forms Generating unit-level electricity output and boiler-level coal input data comes from the 2009-2012 EIA-923 forms. Table 4 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

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

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.[55] 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 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.11 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 and the data source

[54] Available: https://my.vanderbilt.edu/jeremyatack/data-downloads/

[55] See https://www.eia.gov/tools/faqs/faq.cfm?id=107&t=3 for details.

State	First decade of PUC regulation	Data source
Alabama	1920	Troesken (2006)
Arkansas	1920	State PUC website[56]
Iowa	1970	State PUC website[57]
Illinois	1920	Troesken (2006)
Indiana	1920	Troesken (2006)
Kentucky	1940	State PUC website[58]
Minnesota	1980	State PUC website[59]
Missouri	1920	Troesken (2006)
Mississippi	1960	State PUC website[60]
Tennessee	1920	Troesken (2006)
Wisconsin	1910	Troesken (2006)

U.S. 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.[61] This data is used for the estimates in column 4 of Table A.11.

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.[62] These variables are used as outcomes in the regressions shown in columns 1 and 2 of Table A.12.

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.12.

[56] Available: http://www.apscservices.info/commission-history.asp

[57] Available: https://iub.iowa.gov/history

[58] Available: https://psc.ky.gov/Home/About#AbtComm

[59] Available: https://mn.gov/puc/about-us/

[60] Available: https://www.psc.state.ms.us/executive/pdfs/2010/2010%20ANNUAL%20REPORT.pdf

[61] The six regulated criteria pollutants are sulfur dioxide, particulates, nitrogen dioxide, carbon monoxide, ozone, and lead.

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

A.5 Other

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

National 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.2) and total electricity capacity, from fossil fuel and hydropower during 1920-1970 (shown in Figure A.9).[63] Figure A.7 plots the transport cost share of national delivered coal prices during 1902-2007 obtained from McNerney, Farmer and Trancik (2011). Figure A.15 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.[64]

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.

[63] 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.

[64] Available: hathitrust.org

Availability Summaries of these censuses at aggregate data levels are available online.[65] However, power plant-level data could not be located following extensive private conversations with archivists at the National Archives and Records Administration.[66]

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".[67] 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.[68]

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[69]. 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.

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

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

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

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

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

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 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.[70]

C 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. Suppose a county's non-coal electricity only comes from less than 1 MW 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 censoring 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 with 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.

In the absence of historical local data, one can predict the frequency of power plants below the 1MW threshold at the national level and then allocate these missing power plants across electricity-producing counties. Specifically, for fuel j, county i, decade t, denote the capacity of a new power plant indexed by p_{jit} as $X_{p_{jit}}$. 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 $f_{jit}^{0.5}$.

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

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.

Figure A.11 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 unadjusted 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. As expected, the imputation procedure increases the sample size, raises the mean, and reduces the skewness of relative coal capital. In particular, the sample size increases because observations previously with missing relative coal capital values (i.e., positive coal capital and zero non-coal capital) and observations previously with zero relative coal capital values (i.e., zero coal capital and positive non-coal capital) now have strictly positive values. Column 5 displays statistics for relative coal capital using an alternative, less data-driven, imputation which simply adds 1 MW to new coal and non-coal capital investment in 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 county-by-decade observation.

To incorporate estimated uncertainty from the imputation procedure, I iterate Steps 1-4 by bootstrapping the estimation of $g_t()$. This recovers an empirical distribution for each coefficient β^τ from equation (4) arising from estimated uncertainty in the imputation procedure. Figure A.13 shows 95% confidence intervals for $\hat{\beta}^\tau$ constructed from the sum of clustered standard errors using point estimates of $g_t()$ and standard errors from the bootstrap procedure on $g_t()$ with 250 draws.

D Theory appendix

D.1 Model

This section solves the model presented in Section 6.1. In period t for each intermediate sector $j \in \{c, n\}$, the myopic electricity producer chooses capital, X_{jt}, and fuel, E_{jt}, for current-vintage generating units and fuel, E_{jt-1}, for past-vintage generating units. "Clay"-like capital for past-vintage generating units, X_{jt-1}, is fixed. The producer's problem is

$$\max_{X_{ct},X_{nt},E_{ct},E_{nt},E_{ct-1},E_{nt-1}} Y_t - z_{ct}(E_{ct} + E_{ct-1}) - z_{nt}(E_{nt} + E_{nt-1}) - r_t(X_{ct} + X_{nt}) \tag{A.1}$$

where z_{jt} is the fuel price and r_t is the price of capital. Final good Y_t is given by equation (5) and intermediate good Y_{jt} is given by equation (6). To explore how scale and productivity channels could generate path dependence for otherwise similar intermediate sectors, suppose productivities are the same across intermediate goods during period $t-1$, $A_{Xct-1} = A_{Xnt-1}$ and $A_{Ect-1} = A_{Ent-1}$.

Because fuel and capital are perfect complements, efficiency for current-vintage generating units implies that the producer need only to choose capital, and not fuel, set at $A_{Xjt}X_{jt} = A_{Ejt}E_{jt}$ for $j \in \{c, n\}$. Clay-like past-vintage capital further implies $A_{Ejt-1}E_{jt-1} = A_{Xjt-1}(1-\delta)X_{jt-1}$. The resulting constrained optimization problem is

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

with optimality condition

$$\alpha Y_t^{1/\epsilon}(A_{Xjt}A_{Xjt-1}X_{jt-1})^{\frac{\alpha(\epsilon-1)}{\epsilon}} X_{jt}^{\frac{\varphi-1}{\epsilon}} = \frac{A_{Xjt}}{A_{Ejt}}z_{jt} + r_t \quad for \quad j \in \{c, n\} \tag{A.2}$$

where $\varphi = (1-\alpha)(1-\epsilon)$. Taking the ratio of equation (A.2) for coal and non-coal subsectors, and rewriting in terms of current-vintage relative coal capital investment, $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$, yields

$$\widetilde{X}_t = \widetilde{w}_t^{\frac{\epsilon}{\varphi-1}} \widetilde{X}_{t-1}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \widetilde{A}_{Xt}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \tag{A.3}$$

where $w_{jt} = \frac{A_{Xjt}}{A_{Ejt}}z_{jt} + r_t$ is the productivity-weighted input price index, $\widetilde{w} = \frac{w_{ct}}{w_{nt}}$ is relative input price, and $\widetilde{A}_{Xt} = \frac{A_{Xct}}{A_{Xnt}}$ is relative capital productivity for generating units of vintage t. Equation (A.3) is equation (7) in the main text.

D.2 Extension: productivity losses and capital under-utilization

Appendix D.1 assumes that past-vintage capital is fully utilized in the current period. This assumption has previously been used in the putty-clay literature to avoid a well-known curse of dimensionality problem in vintaged capital models (Atkeson and Kehoe, 1999). This section shows how this assumption is justified when there are large efficiency losses associated with using electricity capital below its designed capacity, as discussed in F1 of Section 6.1.

Suppose a generating unit could operate past-vintage capital below its designed capacity. Because $A_{Ejt-1}E_{jt-1} > A_{Xjt-1}(1-\delta)X_{jt-1}$ would lead to wasted inputs, the producer now faces the constraint $A_{Ejt-1}E_{jt-1} \leq A_{Xjt-1}(1-\delta)X_{jt-1}$ for $j \in \{c, n\}$. The resulting constrained optimization problem is

$$\max_{X_{ct},X_{nt},E_{ct-1},E_{nt-1}} \left([A_{Xct}X_{ct}A_{Ect-1}E_{ct-1}]^{\alpha(\epsilon-1)/\epsilon} + [A_{Xnt}X_{nt}A_{Ent-1}E_{nt-1}]^{\alpha(\epsilon-1)/\epsilon} \right)^{\epsilon/(\epsilon-1)}$$
$$- z_{ct}(\frac{A_{Xct}}{A_{Ect}}X_{ct} + E_{ct-1}) - z_{nt}(\frac{A_{Xnt}}{A_{Ent}}X_{nt} + E_{nt-1}) - r_t(X_{ct} + X_{nt})$$
$$s.t. \quad A_{Ejt-1}E_{jt-1} \leq A_{Xjt-1}(1-\delta)X_{jt-1} \quad \text{for} \quad j \in \{c, n\}$$

with optimality conditions

$$\alpha Y_t^{1/\epsilon}(A_{Xjt}A_{Ejt-1}E_{jt-1})^{\frac{\alpha(\epsilon-1)}{\epsilon}} X_{jt}^{\frac{\varphi-1}{\epsilon}} = \frac{A_{Xjt}}{A_{Ejt}} z_{jt} + r_t \quad for \quad j \in \{c, n\} \tag{A.4}$$

$$\alpha Y_t^{1/\epsilon}(A_{Xjt}A_{Ejt-1}X_{jt})^{\frac{\alpha(\epsilon-1)}{\epsilon}} E_{jt-1}^{\frac{\varphi-1}{\epsilon}} = z_{jt} + \lambda_j A_{Ejt-1} \quad for \quad j \in \{c, n\} \tag{A.5}$$

$$\lambda_j[A_{Ejt-1}E_{jt-1} - A_{Xjt-1}(1-\delta)X_{jt-1}] = 0 \quad for \quad j \in \{c, n\} \tag{A.6}$$

$$\lambda_j \geq 0 \quad for \quad j \in \{c, n\} \tag{A.7}$$

$$A_{Ejt-1}E_{jt-1} \leq A_{Xjt-1}(1-\delta)X_{jt-1} \quad for \quad j \in \{c, n\} \tag{A.8}$$

where λ_j is the Lagrange multiplier on each constraint.

Under what conditions would past-vintage capital be underutilized? To answer this, consider each sector j in isolation and take the ratio of equation (A.4) over (A.5). When the constraint binds (i.e., $\lambda_j > 0$) and past-vintage capital is fully utilized, the ratio can be rewritten as

$$\frac{z_{jt}}{w_{jt}} = \frac{X_{jt}}{E_{jt-1}} - \lambda_j \frac{A_{Ejt-1}}{w_{jt}} \tag{A.9}$$

When the constraint does not bind (i.e., $\lambda_j = 0$), this expression becomes

$$\frac{z_{jt}}{w_{jt}} = \frac{X_{jt}^*}{E_{jt-1}^*} \tag{A.10}$$

where the asterisk indicates the non-binding equilibrium. Setting equation (A.9) to equation (A.10) and substituting in the Leontief equality $E_{jt-1}^* = \frac{A_{X^*jt-1}}{A_{E^*jt-1}}(1-\delta)X_{jt-1}^*$ yields

$$A_{xjt-1}\frac{\lambda_j}{w_{jt}} = \frac{1}{1-\delta}\left(\frac{X_{jt}}{X_{jt-1}} - \frac{X_{jt}^*}{X_{jt-1}^*}\frac{A_{E^*jt-1}/A_{X^*jt-1}}{A_{Ejt-1}/A_{Xjt-1}}\right) \tag{A.11}$$

Next, define the fraction of under-utilization of past-vintage capital in the non-binding base as $\varrho_j = \frac{X_{jt-1}^*}{X_{jt-1}} \in [0, 1]$. Further, denote the ratio of current-vintage capital for the non-binding over the binding case as $\varsigma_j = \frac{X_{jt}^*}{X_{jt}} > 0$. Inserting into equation (A.11)

$$A_{xjt-1}\frac{\lambda_j}{w_{jt}} = \frac{1}{1-\delta}\left(\frac{X_{jt}}{X_{jt-1}}\right)\left(1 - \frac{\varsigma_j}{\varrho_j}\frac{A_{E^*jt-1}/A_{X^*jt-1}}{A_{Ejt-1}/A_{Xjt-1}}\right)$$

which implies $\lambda_j > 0$ if and only if

$$\frac{\varrho_j}{\varsigma_j} > \frac{A_{E^*jt-1}/A_{X^*jt-1}}{A_{Ejt-1}/A_{Xjt-1}} > 0 \tag{A.12}$$

The right hand side of Equation (A.12) captures the lost in fuel productivity when past-vintage capital operates below its designed capacity. Vintage capital is more likely to be fully utilized when fuel productivity losses are large.

D.3 Recovering the scale parameter

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

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

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

$$\min_{E_{pt}} \quad \left(z_{pt} + \frac{A_{Ept}}{A_{Xpt}}r_t\right)E_{pt} + z_{pt}Y_{pt}^{1/\alpha}\left(A_{Ept}A_{Ept-1}E_{pt}\right)^{-1} \tag{A.13}$$

Taking the first order condition of equation (A.13) yields a conditional demand function

$$E_{pt}^* = (Y_{pt})^{1/2\alpha}\left(\frac{z_{pt}}{z_{pt} + \frac{A_{Ept}}{A_{Xpt}}r_t}\right)^{1/2}\left(A_{Ept}A_{Ept-1}\right)^{-1/2} \tag{A.14}$$

Inserting equation (A.14) into non-fuel cost at the cost-minimizing input level, 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

$$\ln \text{non\_fuel\_cost}_{pt} = \frac{1}{\psi}\ln Y_{pt} + \frac{1}{2}\ln\left(\frac{z_{pt}}{z_{pt} + \frac{A_{Ept}}{A_{Xpt}}r_t}\right) + \ln r_t + \ln\left(A_{Ept}^{1/2}A_{Ept-1}^{-1/2}A_{Xpt}^{-1}\right) \tag{A.15}$$

where $\psi = 2\alpha$. Equation (A.15) is the structural analog to the OLS specification in equation (8) from the main text with one exception. For ease of exposition, labor is omitted as a factor of production in equation (6) and hence missing from equation (A.15). In practice, the UDI measure of non-fuel cost modeled in regression equation (8) includes labor costs.

D.4 Extension: Convex adjustment costs

This section extends the baseline model by introducing convex adjustment costs in the level of new-vintage capital of the form $\frac{\varsigma}{2}X_{jt}^2$, where $\varsigma \geq 0$ governs the magnitude of these convex capital frictions. The optimality condition of equation (A.2) now becomes

$$\alpha Y_t^{1/\epsilon}(A_{Xjt}A_{Xjt-1}X_{jt-1})^{\frac{\alpha(\epsilon-1)}{\epsilon}} X_{jt}^{\frac{\varphi-1}{\epsilon}} = w_{jt} + \varsigma X_{jt} \quad for \quad j \in \{c,n\} \tag{A.16}$$

where, as in the baseline model, $w_{jt} = \frac{A_{Xjt}}{A_{Ejt}} z_{jt} + r_t$. Taking the ratio of equation (A.16) for coal and non-coal sectors and defining relative coal capital investment as $\widetilde{X}_t = \frac{X_{ct}}{X_{nt}}$ yields

$$\widetilde{X}_t = \left(\frac{w_{ct} + \varsigma X_{ct}}{w_{nt} + \varsigma X_{nt}}\right)^{\frac{\epsilon}{\varphi-1}} \widetilde{X}_{t-1}^{\frac{\alpha(1-\epsilon)}{\varphi-1}} \widetilde{A}_{Xt}^{\frac{\alpha(1-\epsilon)}{\varphi-1}}, \tag{A.17}$$

where $\widetilde{A}_{Xt} = \frac{A_{Xct}}{A_{Xnt}}$ as before.

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 4.

E.1 Parameters

- Reduced-form path dependence parameter: $\widehat{\rho} = \frac{\frac{1}{5}\sum_{\tau=1}^{5}\hat{\beta}^{\tau}}{\hat{\pi}} = 1.45$, with standard error $\widehat{\sigma}_\rho = 0.72$ (from column 2 of Table 2)
- Returns to scale: $\widehat{\psi} = 1.66$, with standard error $\widehat{\sigma}_\psi = 0.24$ (from column 2 of Table 3)
- 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.15)
- 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[71]
- Carbon content of natural gas: $C_n = 119.9$ lb CO_2/thousand cubic feet[72]

[71] Available here: https://www.eia.gov/environment/emissions/co2_vol_mass.cfm
[72] Available here: https://www.eia.gov/environment/emissions/co2_vol_mass.cfm

E.2 Historical emissions

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

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

where E_{ct} and E_{nt} is U.S. average county coal (in short tons) and natural gas (in thousand cubic feet) consumed, respectively, by the electricity sector in decade t.[73]

E.3 Simulating future emissions

For each combination of shock duration, $d \in \{10, 30, 50\}$, and shock multiplier, $M \in \{1, 2, 6\}$, 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

- Draw $\psi(b) \sim N(\widehat{\psi}, \widehat{\sigma_\psi})$ and $\rho(b) \sim N(\widehat{\rho}, \widehat{\sigma}_\rho)$. Define $\alpha(b) = \frac{\psi(b)}{2}$
- Obtain $\epsilon(b) = 1 + \frac{\rho(b)}{\frac{\psi(b)}{2} - \rho(b)\left(1 - \frac{\psi(b)}{2}\right)}$ and 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

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

[73] E_{ct} and E_{nt} obtained from the Energy Information Administration (2012).

5. Obtain natural gas capital

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

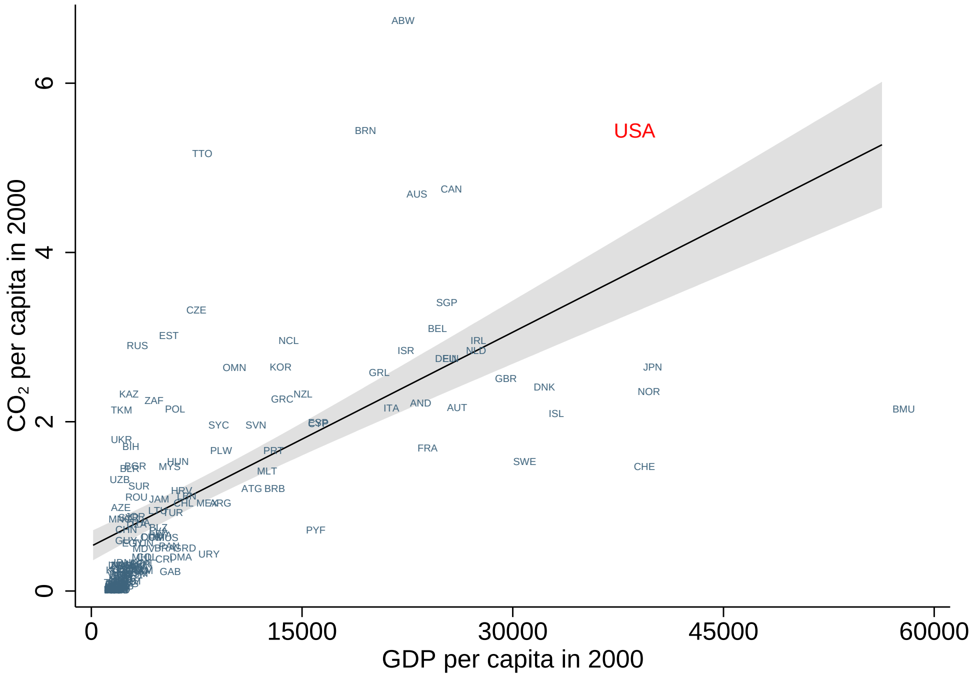
6. Obtain total CO_2 emissions using 2000 emissions intensity

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

Figure 4 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, $d \in \{10, 30, 50\}$, and shock multiplier, $M \in \{1, 2, 6\}$. 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 5 replicates the above simulation procedure for different values of ϵ and ψ with the shock multiplier set at $M = 1$. The heat map plotted in Figure 5 shows the minimum shock duration d needed for long-term CO_2 emissions to be weakly declining.

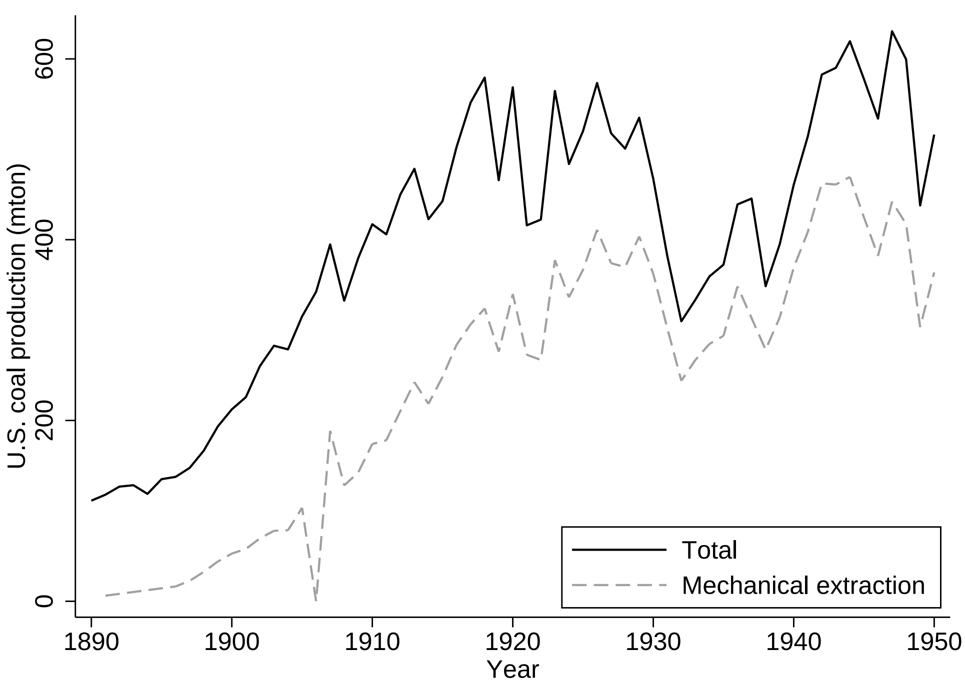
Appendix Figures

Figure A.1: CO_2 emissions intensity and income in 2000



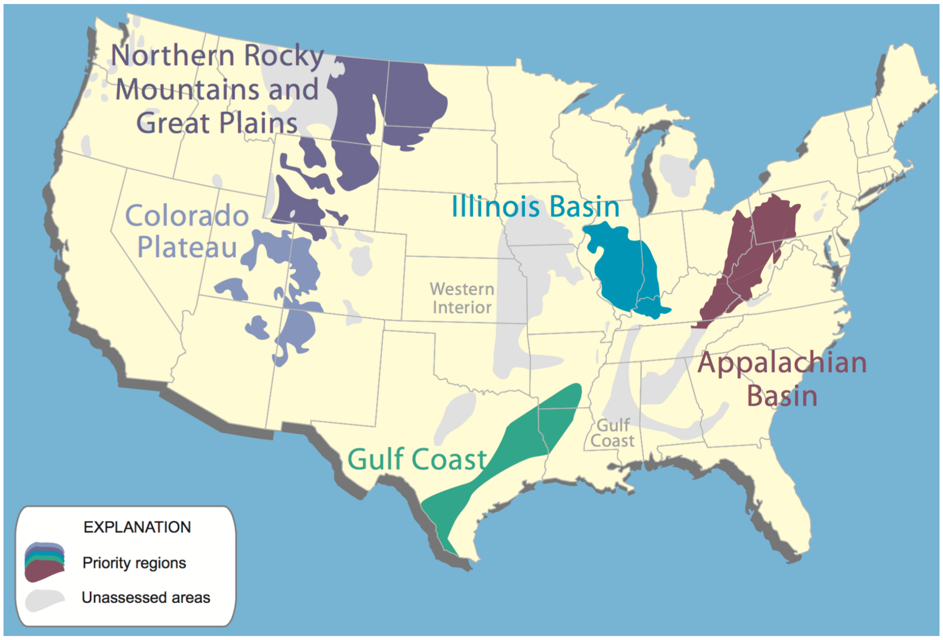
NOTES: Plot shows carbon dioxide emissions (in tons of carbon) per capita against GDP (in nominal USD) per capita in 2000. Linear regression fit shown with 95% confidence interval. OPEC countries excluded. Data from Boden, Marland and Andres (2013) and Bank (2014).

Figure A.2: U.S. bituminous coal production and mechanization



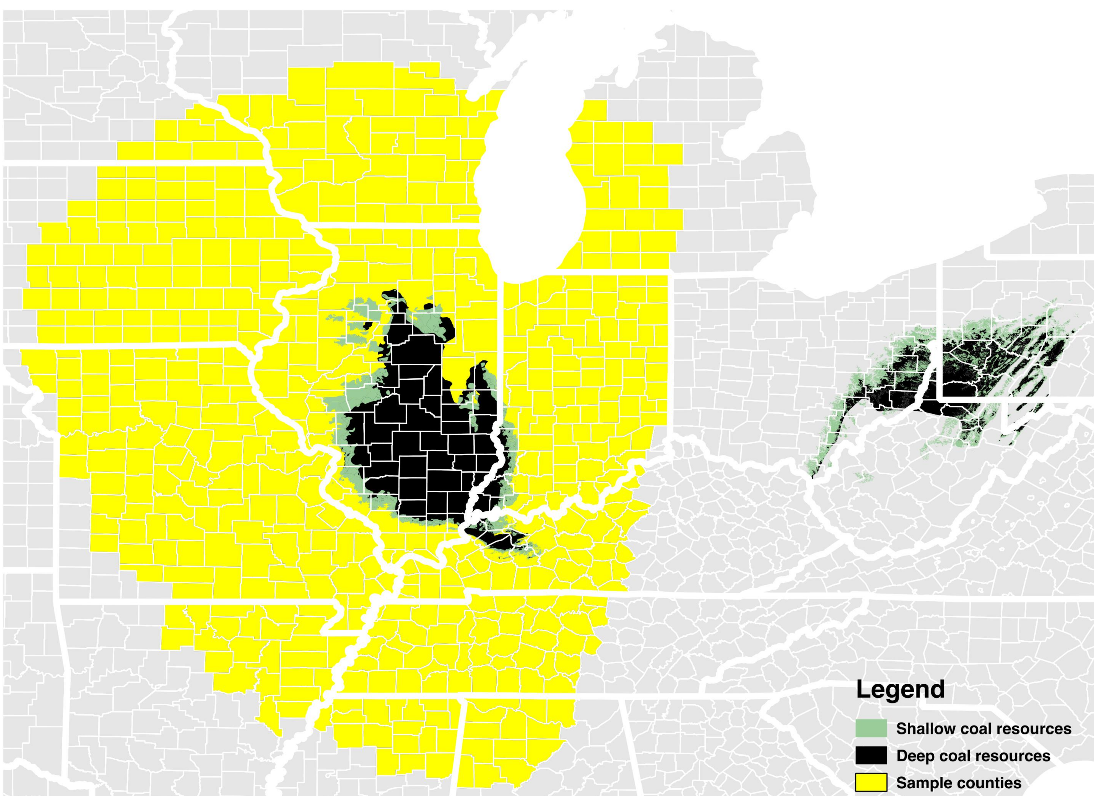
NOTES: Annual U.S. bituminous coal production (in mega short tons). Solid black line shows total coal production. Dashed gray line shows coal production from mechanical extraction. Data from the U.S. Census Bureau (1975).

Figure A.3: U.S. coal basins



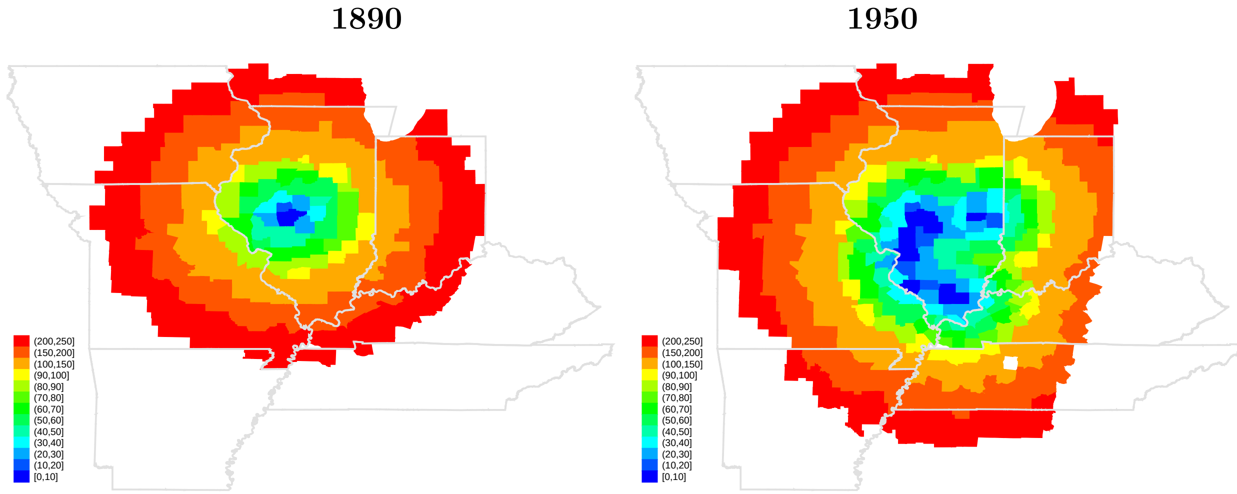
NOTES: Map of major U.S. coal basins. Reproduced from East (2012).

Figure A.4: Location of sample counties and coal basins



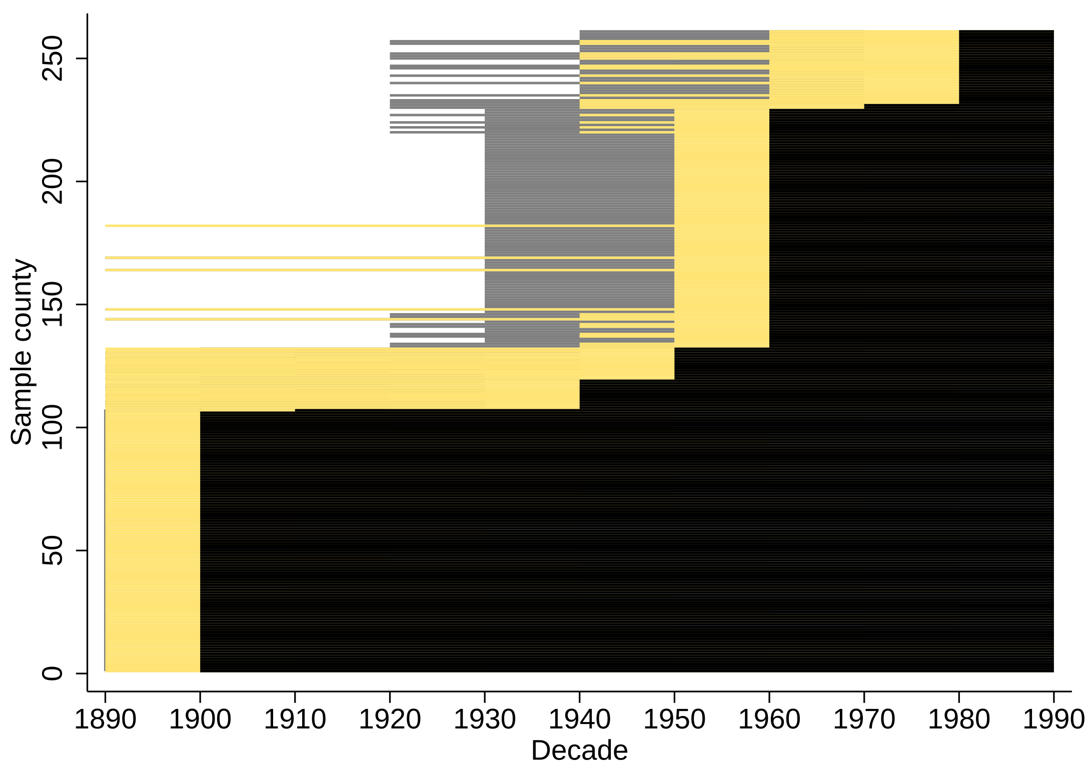
NOTES: A county is included in the baseline sample (in yellow shading) if its spatial centroid is (i) closer to coal resources in the Illinois Basin than in the Appalachian Basin and (ii) less than 250 miles from nearest Illinois coal resource. Shallow (< 200 ft. underground) and deep (> 200 ft. underground)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

Figure A.5: County distance to nearest mine in 1890 and 1950



NOTES: County distance to nearest coal mine in 1890 and 1950 over sample counties.

Figure A.6: Timing of shallow to deep coal switching for each sample county

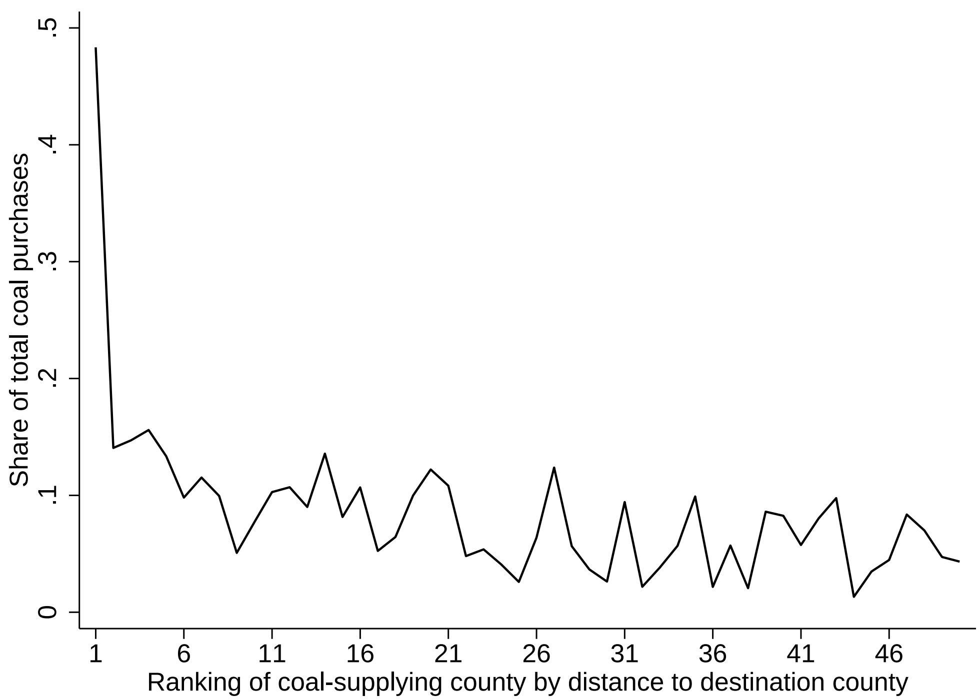


NOTES: The timing of when a sample county's nearest mine switches from a shallow to a deep coal mine for the first time. Counties are stacked according to the decade when the switching event occurs. The gray, yellow, and black shaded areas correspond to event-time periods $h < 0$, $h = 0$, and $h > 0$, respectively. $h = 0$ can span multiple decades if there are several decades between the initial switch to a deep coal mine and the previous switch in coal supplier.

Figure A.7: Share of delivered coal price due to transport costs at the national level

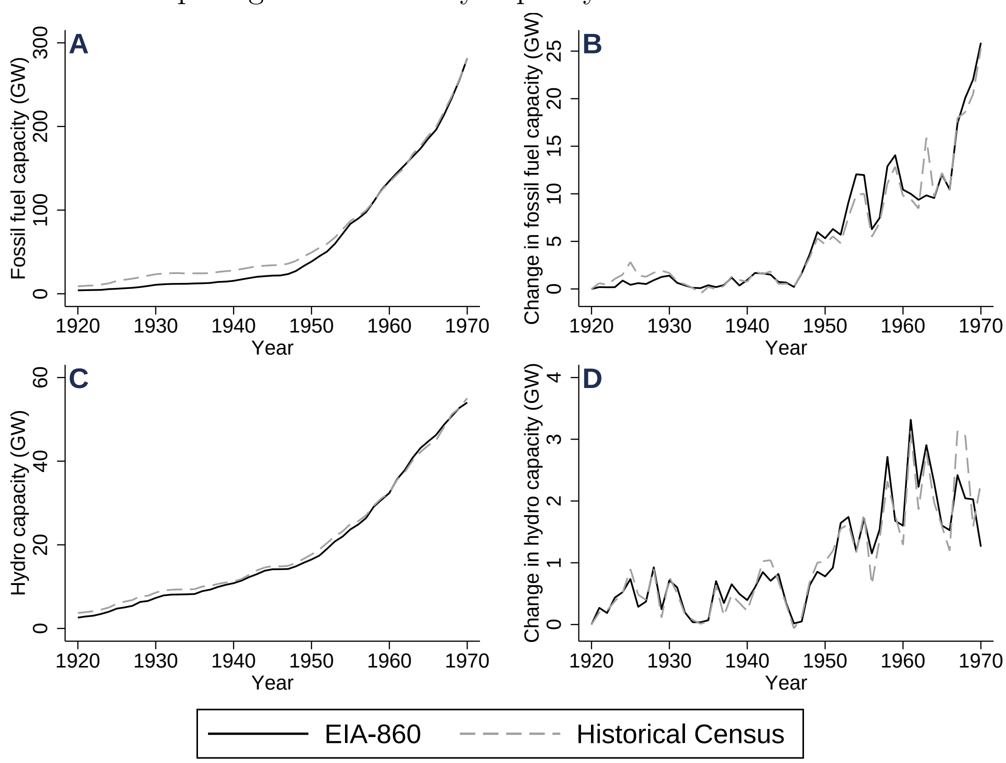
NOTES: Share of transport cost (nominal USD per short ton) in delivered coal price (nominal USD per short ton) for the U.S. during 1902-2007. Data from McNerney, Farmer and Trancik (2011).

Figure A.8: Testing the Herfindahl Principle



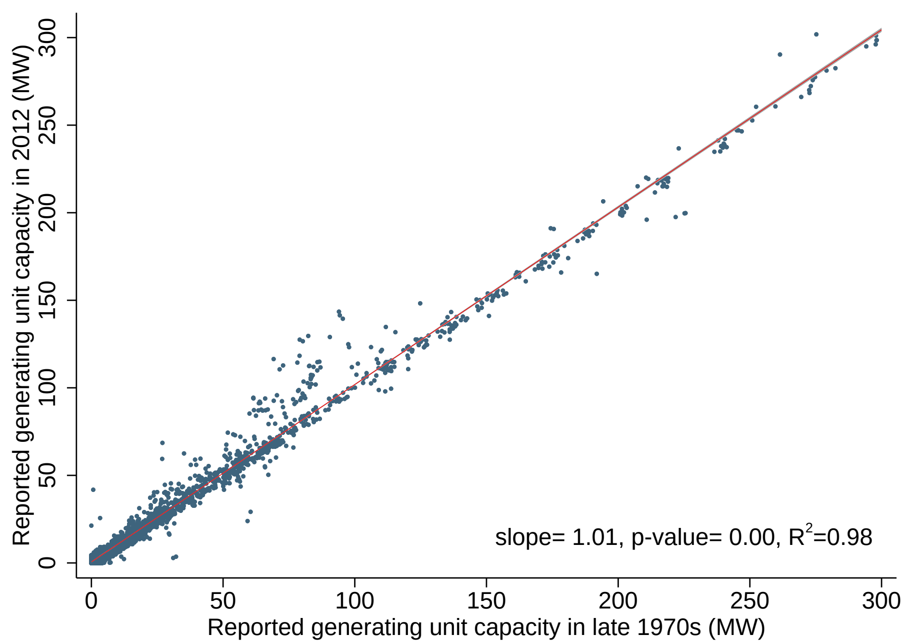
NOTES: Vertical axis shows share of total coal purchase from destination county. Horizontal axis shows the ranking of bilateral distance between spatial centroids of origin and destination counties. Data averaged over 1990-1999 and all U.S. counties that purchase coal for electricity.

Figure A.9: Comparing U.S. electricity capacity from EIA-860 vs. historical census



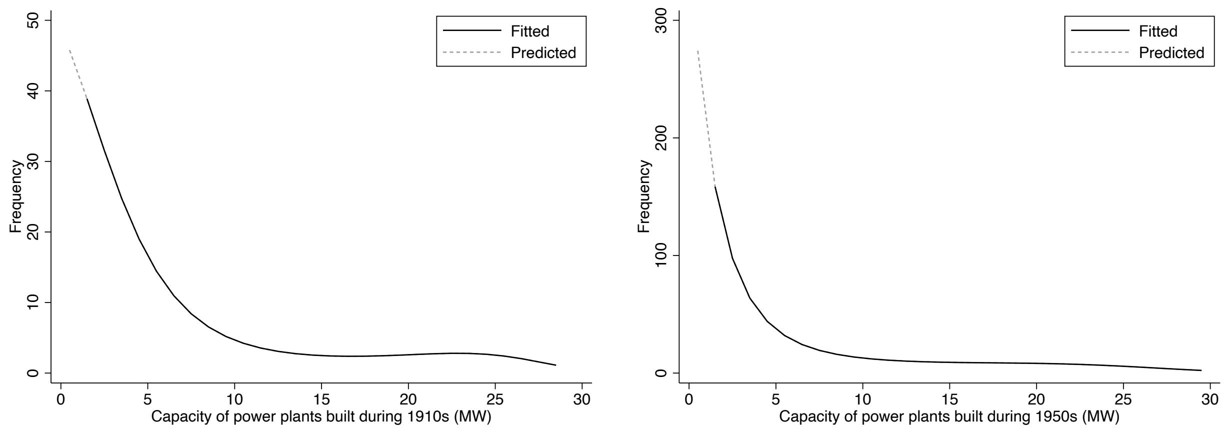
Notes: Panel (A) compares annual aggregate U.S. electricity capacity (in GW) using fossil fuels constructed from EIA-860 forms (solid black lines) against values from the U.S. Historical Census (dashed gray lines) over 1920-1970. Panel (B) plots capacity changes. Panels (C) and (D) show the same information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

Figure A.10: Comparing generating unit capacity in late 1970s and 2012



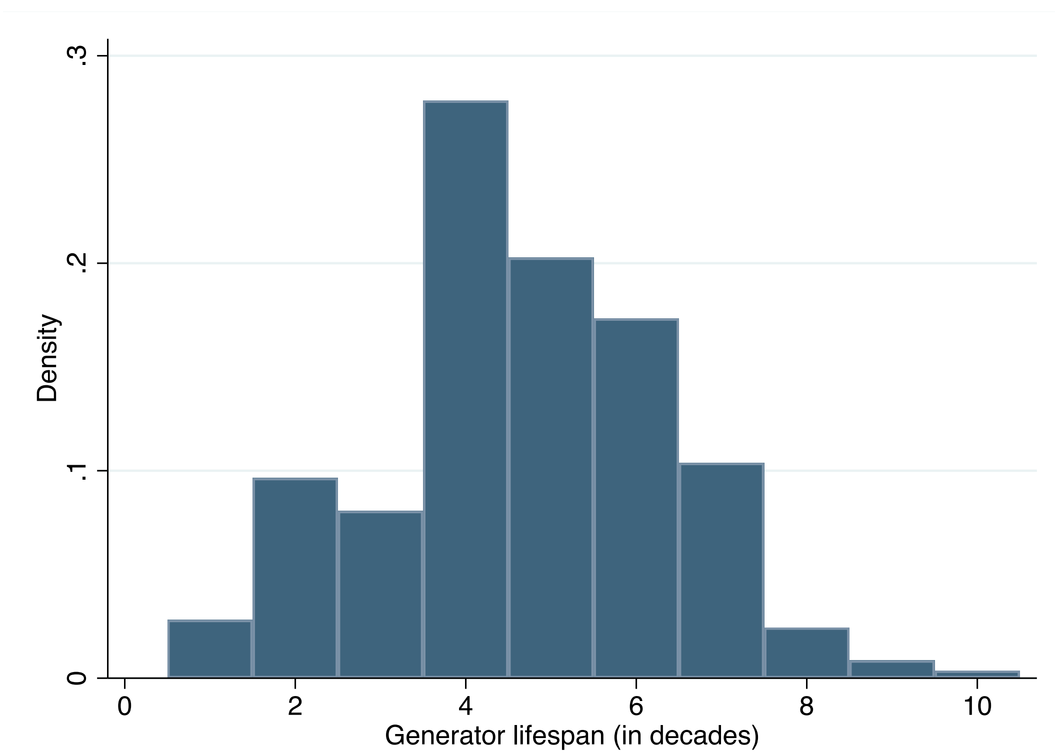
Notes: Scatter shows reported generating unit capacity (in MW) in the 2012 EIA-860 form against reported generating unit capacity in the late 1970s (in MW). Both axes are truncated at 300 MW. Linear fit in red with 95% confidence interval shown in gray. Coefficient, p-value, and R^2 shown from a linear regression using all matched generating units.

Figure A.11: Fitted and predicted capacity distribution of power plants by decade of opening



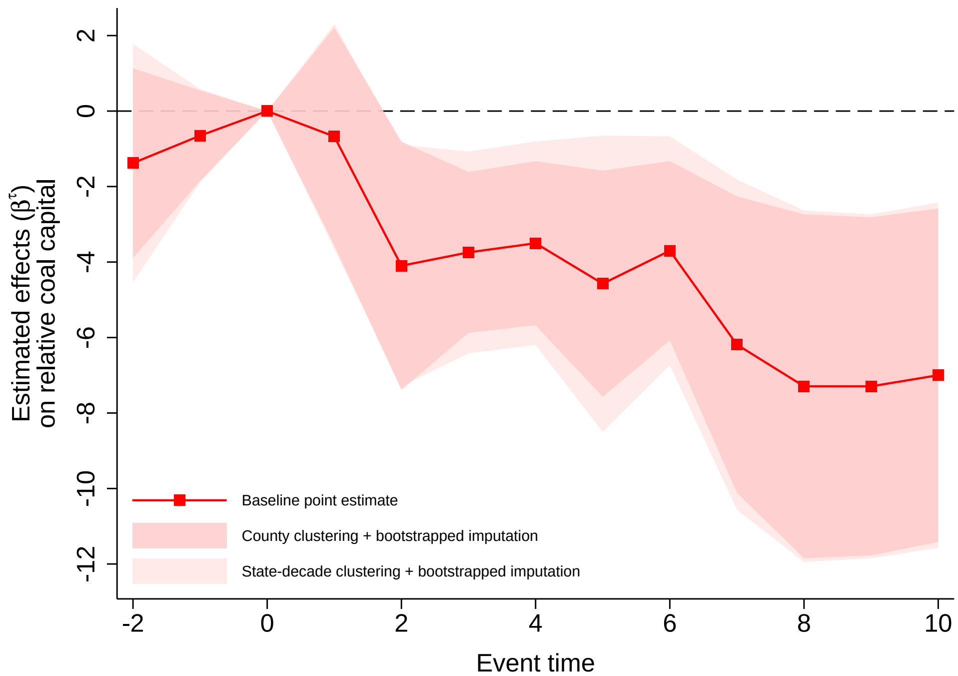
NOTES: Solid line shows fitted capital frequency distribution of all power plants built during 1910s (left panel) and 1950s (right panel). Fitted relationship uses a 4th order polynomial function for power plants with capacity less than 30 MW and greater than 1 MW. Dashed line shows predicted capacity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

Figure A.12: Distribution of sample generator unit lifespans



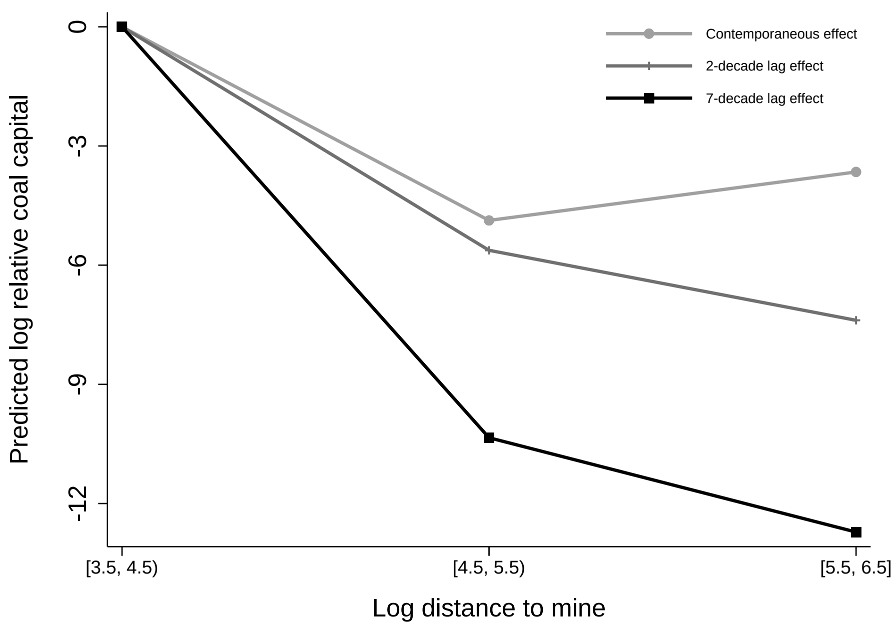
NOTES: Distribution of lifespans for sample generating units.

Figure A.13: Reduced-form estimates of path dependence: bootstrapped imputation



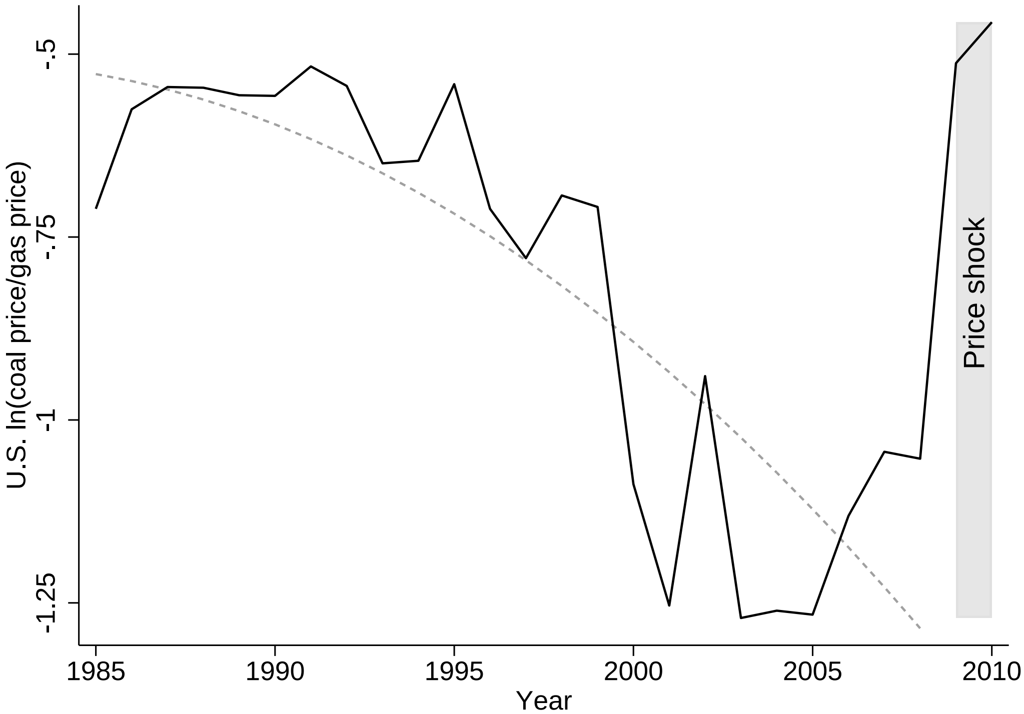
Notes: Thick solid red line show baseline point estimates and 95% confidence intervals for β^τ from equation (4). Outcome is relative coal capital at the county-by-decade level. County sample shown in Figure A.4. Time period is 1890-1990. Darker (lighter) shaded area shows 95% confidence interval with county (state-by-decade) level clustered standard errors and estimated imputation uncertainty from bootstrapping the imputation procedure for missing small power plants with 250 draws. See Appendix C for bootstrap details.

Figure A.14: 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.15: 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 ln delivered coal price		
$\ln d_{it}$	0.38**	0.58***	0.34***
	(0.15)	(0.15)	(0.11)
Decade	1970s	1980s	1990s
Counties	153	153	133

NOTES: Each column is a separate cross-sectional regression of log delivered coal price (in nom. USD per ton) averaged within each decade on log distance to nearest mine and state fixed effects. County sample shown in Figure A.4. 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

Number of different values	Percentage of generating units with different reported 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 2012				
Primary fuel in 1970s	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 unadjusted and imputed relative coal capital

	(1) Unadjusted	(2) Imputed 3rd order poly.	(3) Imputed 4th order poly.	(4) Imputed 5th order poly.	(5) Imputed add 1 MW
			Number of observations		
Total	2,369	2,369	2,369	2,369	2,369
Missing	1,246	0	0	0	0
Zero	825	0	0	0	0
Positive	298	2,369	2,369	2,369	2,369
			Summary statistics		
Obs	1,123	2,369	2,369	2,369	2,369
Mean	9.71	65.01	75.58	65.14	8.54
Median	0	.15	.16	.14	1
SD	53.44	355.37	417.01	356.52	37.3
Skewness	11.51	7.39	7.47	7.41	6.76

NOTES: Top panel shows the number of total, missing, zero-valued, and positive-valued observations for the baseline county-by-decade sample shown in Figure A.4. Bottom panel shows various summary statistics. Column 1 shows unadjusted relative coal capital. Columns 2-4 adds imputed missing power plants with capacity less than 1 MW using 3rd, 4th, and 5th order polynomial functions for $g_t()$, respectively. Column 5 adjusts relative coal capital by adding 1 MW to both coal and non-coal capital investment. See Appendix C for details.

Table A.6: Robustness: additional identification concerns

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome is relative coal capital					
$\ln d_i^0$ (β^τ)						
2 decades lead	-1.38	-1.49	-0.71	-2.10	-0.64	-1.41
	(1.02)	(0.96)	(1.62)	(1.91)	(1.09)	(0.92)
1 decade lead	-0.65	-1.14	-0.22	-1.29	-1.04	0.40
	(0.67)	(0.71)	(0.85)	(0.82)	(0.99)	(0.98)
	—	—	—	—	—	—
1 decade lag	-0.67	-2.22**	-0.72	-1.07	-0.15	-0.71
	(1.25)	(1.01)	(1.05)	(1.40)	(1.10)	(1.30)
2 decades lag	-4.10***	-4.91***	-3.19***	-5.03***	-3.92***	-4.18***
	(1.14)	(1.39)	(1.08)	(1.20)	(0.86)	(1.19)
3 decades lag	-3.74***	-4.28***	-4.10***	-4.87***	-4.09***	-3.73***
	(0.66)	(0.62)	(1.39)	(1.10)	(0.90)	(0.73)
4 decades lag	-3.50***	-4.32***	-4.10***	-4.86***	-4.13***	-3.51***
	(0.71)	(0.67)	(1.53)	(1.22)	(0.99)	(0.75)
5 decades lag	-4.57***	-5.40***	-5.95***	-5.53***	-5.19***	-4.59***
	(0.98)	(1.09)	(1.69)	(1.15)	(1.20)	(0.99)
6 decades lag	-3.70***	-4.52***	-4.83***	-4.84***	-3.94***	-3.72***
	(0.75)	(0.67)	(1.43)	(1.15)	(0.89)	(0.79)
7 decades lag	-6.19***	-7.11***	-6.99***	-8.10***	-6.13***	-6.21***
	(1.37)	(1.45)	(1.81)	(1.52)	(1.26)	(1.39)
8 decades lag	-7.29***	-8.26***	-8.03***	-9.52***	-7.13***	-7.31***
	(1.59)	(1.72)	(1.93)	(1.68)	(1.46)	(1.61)
9 decades lag	-7.29***	-8.22***	-8.02***	-10.26***	-7.19***	-7.32***
	(1.57)	(1.69)	(2.05)	(1.69)	(1.54)	(1.58)
10 decades lag	-7.00***	-7.85***	-7.95***	-10.63***	-7.31***	-7.02***
	(1.56)	(1.68)	(2.06)	(1.84)	(1.48)	(1.58)
$\ln d_{it}$ (π)	-1.53***	-1.69***	-1.56**	-1.86***	-1.37**	-1.53***
	(0.53)	(0.55)	(0.62)	(0.50)	(0.59)	(0.53)
Dropped if mine becomes 2nd closest	No	Yes	No	No	No	No
Demand controls	No	No	Yes	No	No	No
Geographic controls	No	No	No	Yes	No	No
Wage control	No	No	No	No	Yes	No
Single decade period before switch	No	No	No	No	No	Yes
Observations	2369	2033	2106	2369	2097	2230
Counties	261	219	261	261	261	261

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.4. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates. Column 2 estimates baseline model but drops counties for which the shallow mine becomes the second nearest mine in any decade after the switching event. Column 3 adds county-by-decade population, number of manufacturing establishments, and manufacturing employment, all in logs. Column 4 adds log county distance to nearest navigable waterway and log variance in slope, both interacted with a linear time trend. Column 5 adds county-by-decade log manufacturing wages. Column 6 redefines the focal event as just the single decade before the initial switch to deep coal. 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)	(6)	(7)	(8)
$\ln d_i^0$ (β^τ)								
2 decades lead	-1.38	-1.30	-1.33	-0.63	-0.23	-2.66	-0.10	-0.07
	(1.02)	(1.01)	(1.00)	(0.96)	(0.46)	(2.24)	(1.71)	(0.75)
1 decade lead	-0.65	-0.70	-0.68	-0.45	0.01	-1.53	1.14	0.16
	(0.67)	(0.68)	(0.67)	(0.60)	(0.31)	(0.95)	(1.61)	(0.29)
	—	—	—	—	—	—	—	—
1 decade lag	-0.67	-0.58	-0.59	-0.51	-0.12	-0.84	-0.99	-0.21*
	(1.25)	(1.20)	(1.19)	(1.07)	(0.33)	(0.59)	(0.81)	(0.12)
2 decades lag	-4.10***	-3.97***	-3.97***	-4.27***	-1.90***	-2.49	-2.88**	-0.47**
	(1.14)	(1.08)	(1.08)	(1.14)	(0.44)	(2.04)	(1.14)	(0.21)
3 decades lag	-3.74***	-3.70***	-3.68***	-4.11***	-2.11***	-4.18**	-3.77**	-0.62**
	(0.66)	(0.66)	(0.66)	(0.89)	(0.41)	(2.04)	(1.63)	(0.27)
4 decades lag	-3.50***	-3.44***	-3.43***	-3.93***	-2.03***	-4.97**	-4.25***	-0.69**
	(0.71)	(0.71)	(0.71)	(0.96)	(0.37)	(1.96)	(1.62)	(0.31)
5 decades lag	-4.57***	-4.49***	-4.46***	-5.02***	-2.29***	-3.91*	-3.16*	-0.97***
	(0.98)	(0.91)	(0.91)	(1.19)	(0.37)	(2.09)	(1.69)	(0.31)
6 decades lag	-3.70***	-3.66***	-3.63***	-4.11***	-1.47***	-2.82	-2.95*	-0.91***
	(0.75)	(0.75)	(0.75)	(0.98)	(0.48)	(2.21)	(1.73)	(0.33)
7 decades lag	-6.19***	-6.15***	-6.12***	-6.64***	-3.18***	0.71	-1.29	-0.66
	(1.37)	(1.37)	(1.36)	(1.54)	(0.80)	(2.99)	(3.00)	(0.40)
8 decades lag	-7.29***	-7.24***	-7.22***	-7.91***	-3.90***	-3.33	-1.94	-0.66
	(1.59)	(1.58)	(1.58)	(1.79)	(0.80)	(2.66)	(2.07)	(0.47)
9 decades lag	-7.29***	-7.25***	-7.22***	-7.87***	-3.95***	-2.63	-1.27	-0.53
	(1.57)	(1.56)	(1.55)	(1.76)	(0.78)	(2.47)	(1.95)	(0.47)
10 decades lag	-7.00***	-6.96***	-6.93***	-7.58***	-3.76***	-2.20	-0.79	-0.55
	(1.56)	(1.55)	(1.55)	(1.76)	(0.75)	(2.85)	(2.30)	(0.51)
$\ln d_{it}$ (π)	-1.53***	-1.54***	-1.54***	-1.59***	-0.86**	-2.41***	-1.03	-0.16
	(0.53)	(0.53)	(0.53)	(0.55)	(0.40)	(0.75)	(1.10)	(0.15)
Outcome	relative coal	relative coal	relative coal	relative coal	relative coal	relative coal	relative coal	coal share
Small plant imputation	4^{th} ord. poly	3^{rd} ord. poly	5^{th} ord. poly	4^{th} ord. poly	add 1MW	none	1-2MW	none
Weights	none	none	none	plants	none	none	none	none
Observations	2369	2369	2369	2631	2369	369	369	565
Counties	261	261	261	261	261	65	65	97

NOTES: Estimates of β^τ and π from equation (4) using a Poisson model. County sample shown in Figure A.4. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Outcome variable in columns 1-5 is relative coal capital at the county-by-decade level. Column 1 replicates baseline estimates which imputes missing small power plants with a 4th order polynomial function for $g_t()$ to 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 weights each observation using a county's time-averaged number of power plants. Outcome in column 5 adds 1 MW to both unadjusted coal and non-coal capital investment to construct relative coal capital. Column 6 constructs relative coal capital using the unadjusted data. Column 7 pretends 1-2 MW power plants are missing and imputes them using the same procedure used to impute missing $<$ 1 MW plants. Column 8 models coal capital share using the unadjusted data. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.8: Robustness: sample restrictions

	(1)	(2)	(3)	(4)	(5)	(6)
	Outcome is relative coal capital					
$\ln d_i^0$ (β^τ)						
2 decades lead	-1.38	-4.26**	-0.98	-2.22*	-1.20	-1.42
	(1.02)	(1.67)	(1.07)	(1.27)	(1.01)	(1.05)
1 decade lead	-0.65	-2.53**	-0.52	-1.26	-0.49	-0.75
	(0.67)	(1.04)	(0.65)	(0.82)	(0.75)	(0.66)
	—	—	—	—	—	—
1 decade lag	-0.67	-0.93	-0.60	-1.98	-0.23	-0.71
	(1.25)	(1.40)	(1.22)	(1.23)	(1.16)	(1.05)
2 decades lag	-4.10***	-5.32***	-4.01***	-5.99***	-2.70***	-3.93***
	(1.14)	(1.02)	(1.12)	(1.16)	(1.02)	(1.03)
3 decades lag	-3.74***	-5.76***	-3.65***	-4.60***	-2.33***	-3.70***
	(0.66)	(1.14)	(0.65)	(0.83)	(0.90)	(0.65)
4 decades lag	-3.50***	-5.38***	-3.39***	-4.48***	-2.00**	-3.48***
	(0.71)	(1.39)	(0.71)	(0.93)	(0.96)	(0.69)
5 decades lag	-4.57***	-3.75**	-4.45***	-7.28***	-1.69	-4.13***
	(0.98)	(1.62)	(0.95)	(2.07)	(1.12)	(0.77)
6 decades lag	-3.70***	1.89	-3.62***	-4.48***	-1.59	-3.83***
	(0.75)	(3.78)	(0.74)	(0.98)	(1.16)	(0.73)
7 decades lag	-6.19***	-6.10**	-6.16***	-6.90***	-3.71**	-4.55***
	(1.37)	(3.04)	(1.37)	(1.37)	(1.50)	(0.97)
8 decades lag	-7.29***	-7.79**	-7.27***	-8.07***	-4.62***	-5.15***
	(1.59)	(3.08)	(1.59)	(1.54)	(1.69)	(1.21)
9 decades lag	-7.29***	-7.60**	-7.27***	-8.09***	-4.43***	-5.11***
	(1.57)	(3.06)	(1.56)	(1.48)	(1.66)	(1.23)
10 decades lag	-7.00***	-7.87**	-6.98***	-7.75***	-4.18**	-5.05***
	(1.56)	(3.19)	(1.56)	(1.49)	(1.67)	(1.23)
$\ln d_{it}$ (π)	-1.53***	1.87	-1.56***	-1.53***	-1.33**	-0.75
	(0.53)	(1.37)	(0.53)	(0.53)	(0.52)	(0.59)
Sample	Benchmark	$> 90^{th}$ pct mines	$> 97.5^{th}$ pct mines	<200 miles from Ill. coal	<300 milles from Ill. coal	Incl. closer to App. coal
Observations	2369	1920	2402	1927	2869	3205
Counties	261	223	261	207	319	337

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.4. Column 2 uses transport distance constructed from mines with area above the 90th percentile. Column 3 uses transport distance constructed from mines with area above the 97.5th percentile. Columns 4 and 5 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 6 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. *** p<0.01, ** p<0.05, * p<0.1.

Table A.9: Robustness: alternative modeling choices

	(1)	(2)	(3)	(4)	(5)
$\ln d_i^0$ (β^τ)					
3 decades lead			0.87		
			(1.48)		
2 decades lead	-1.38		-1.35	-0.27	0.32
	(1.02)		(1.00)	(0.42)	(1.04)
1 decade lead	-0.65	-0.71	-0.63	0.27	0.27
	(0.67)	(0.69)	(0.67)	(0.40)	(0.64)
	—	—	—	—	—
1 decade lag	-0.67	-0.64	-0.70	-0.41	-0.76*
	(1.25)	(1.23)	(1.27)	(0.26)	(0.39)
2 decades lag	-4.10***	-4.05***	-4.15***	-1.21***	-1.40**
	(1.14)	(1.14)	(1.14)	(0.31)	(0.60)
3 decades lag	-3.74***	-3.69***	-3.81***	-1.71***	-1.93**
	(0.66)	(0.66)	(0.67)	(0.49)	(0.87)
4 decades lag	-3.50***	-3.44***	-3.58***	-1.80***	-2.35**
	(0.71)	(0.72)	(0.72)	(0.55)	(1.04)
5 decades lag	-4.57***	-4.52***	-4.64***	-2.18***	-2.61**
	(0.98)	(0.97)	(0.99)	(0.65)	(1.06)
6 decades lag	-3.70***	-3.64***	-3.77***	-1.79**	-2.49**
	(0.75)	(0.75)	(0.75)	(0.85)	(1.12)
7 decades lag	-6.19***	-6.13***	-6.26***	-2.88**	-3.23
	(1.37)	(1.37)	(1.38)	(1.28)	(2.04)
8 decades lag	-7.29***	-7.23***	-7.36***	-4.10***	-4.19**
	(1.59)	(1.59)	(1.60)	(1.37)	(2.01)
9 decades lag	-7.29***	-7.24***	-7.36***	-3.70**	-3.86*
	(1.57)	(1.56)	(1.57)	(1.51)	(2.03)
10 decades lag	-7.00***	-6.94***	-7.07***	-3.74**	-3.55*
	(1.56)	(1.56)	(1.57)	(1.70)	(2.08)
$\ln d_{it}$ (π)	-1.53***	-1.53***	-1.53***	-0.18	-0.66
	(0.53)	(0.53)	(0.53)	(0.62)	(0.90)
Model	Poisson	Poisson	Poisson	Linear	Neg. bin.
Observations	2369	2240	2498	2369	2369
Counties	261	261	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.4. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline model using a Poisson model with relative coal capital as the outcome and includes 2 lead terms. Column 2 is identical to column 1 except for having 1 lead term. Column 3 is identical to column 1 except for having 3 lead terms. Column 4 uses a log-log linear model with log relative coal capital as the outcome and includes 2 lead terms. Column 5 uses a negative binomial model with dispersion parameter that is a function of expected relative coal capital and includes 2 lead terms. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.10: Robustness: staggered treatment

	(1)	(2)	(3)	(4)
	Outcome is relative coal capital			
$\ln d_i^0$ (β^τ)				
2 decade lead	-1.38	-1.48	-1.52	
	(1.02)	(0.98)	(1.33)	
1 decade lead	-0.65	-1.03	-0.63	0.04
	(0.67)	(0.73)	(0.72)	(0.11)
	—	—	—	—
1 decade lag	-0.67	-2.20**	-1.29	0.82
	(1.25)	(0.99)	(1.34)	(0.54)
2 decade lag	-4.10***	-4.88***	-2.64	-1.72**
	(1.14)	(1.40)	(2.35)	(0.71)
3 decade lag	-3.74***	-4.26***	-3.82***	-2.27**
	(0.66)	(0.61)	(0.73)	(0.90)
4 decade lag	-3.50***	-4.35***	-3.77***	
	(0.71)	(0.65)	(0.92)	
5 decade lag	-4.57***	-5.40***	-5.04***	
	(0.98)	(1.00)	(1.11)	
6 decade lag	-3.70***	-4.42***	-3.69***	
	(0.75)	(0.64)	(0.72)	
7 decade lag	-6.19***	-6.91***	-6.19***	
	(1.37)	(1.30)	(1.35)	
8 decade lag	-7.29***	-8.04***	-7.35***	
	(1.59)	(1.53)	(1.58)	
9 decade lag	-7.29***	-8.06***	-7.31***	
	(1.57)	(1.52)	(1.55)	
10 decade lag	-7.00***	-7.72***	-7.02***	
	(1.56)	(1.50)	(1.53)	
1 decade lag x switch decade			0.09	
			(0.16)	
2 decade lag x switch decade			-0.22	
			(0.24)	
3 decade lag x switch decade			0.09	
			(0.08)	
4 decade lag x switch decade			0.05	
			(0.15)	
5 decade lag x switch decade			0.18	
			(0.17)	
$\ln d_{it}$ (π)	-1.53***	-1.54***	-1.54***	-1.64
	(0.53)	(0.53)	(0.53)	(1.56)
Sample	Full	Pre-1960 event	Full	Stacked
Controls counties	All	All	All	Not-yet/never treated
Observations	2369	2149	2369	869
Counties	261	229	261	132

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.4. Time period is 1890-1990. Each model includes event time, county, and state-by-decade fixed effects. Column 1 replicates baseline estimates. Column 2 restricts sample to counties that experience the switching event prior to the 1960s. Column 3 interacts lagged terms up to five decades with the decade in which a county experiences the switching event. Column 4 employs a stacked difference-in-difference approach following, restricting the sample to counties with balanced event time observations and limiting control counties to not-yet and never treated counties. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.11: Other mechanisms: cost-of-service and Clean Air Act regulations

	(1)	(2)	(3)	(4)
	Outcome is relative coal capital			
$\ln d_i^0$ (β^τ)				
2 decades lead	-1.38	-0.75	-1.62	0.08
	(1.02)	(0.92)	(1.00)	(1.39)
1 decade lead	-0.65	-0.71	-1.46	0.53
	(0.67)	(0.56)	(1.00)	(1.12)
	—	—	—	—
1 decade lag	-0.67	-2.80**	-1.90**	-3.08**
	(1.25)	(1.10)	(0.83)	(1.40)
2 decades lag	-4.10***	-2.93***	0.03	-3.25**
	(1.14)	(1.11)	(0.88)	(1.29)
3 decades lag	-3.74***	-2.92*	-3.11***	-4.13***
	(0.66)	(1.55)	(1.07)	(0.80)
4 decades lag	-3.50***	-6.94*	-2.45	-3.71***
	(0.71)	(3.91)	(2.00)	(1.13)
5 decades lag	-4.57***	-3.10***	-5.51*	-2.26*
	(0.98)	(1.13)	(2.90)	(1.26)
6 decades lag	-3.70***	-3.11***	-4.45	2.09
	(0.75)	(1.14)	(2.93)	(2.19)
7 decades lag	-6.19***		-4.81*	-4.32*
	(1.37)		(2.55)	(2.25)
8 decades lag	-7.29***			-5.77**
	(1.59)			(2.42)
9 decades lag	-7.29***			-5.96***
	(1.57)			(2.14)
10 decades lag	-7.00***			-5.33**
	(1.56)			(2.10)
$\ln d_{it}$ (π)	-1.53***	0.36	0.74	-1.76***
	(0.53)	(1.10)	(1.13)	(0.68)
Keep before PUC?	No	Yes	No	No
Drop ever in nonattainment?	No	No	No	Yes
Sample period	1890-1990	1890-1970	1890-1960	1890-1990
Observations	2369	745	1586	1683
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.4. 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 U.S. Clean Air Act. Robust standard errors clustered at the county level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table A.12: Other mechanisms: upstream and downstream sectors

	(1)	(2)	(3)	(4)
	Outcome is			
	ln railroad density	ln highway density	Env. NGO share	Rpublican vote share
$\ln d_i^o \times sinceEvent_i$ (ω_1)	-0.048	-0.022	-0.000	0.003
	(0.031)	(0.018)	(0.000)	(0.003)
$\ln d_i^o$ (ω_2)	-0.002	0.013	-0.001	-0.002
	(0.169)	(0.114)	(0.000)	(0.021)
$sinceEvent_i$ (ω_3)	0.226*	0.051	-0.000	-0.011
	(0.135)	(0.072)	(0.000)	(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.4. 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.