Bargaining, Merger, Capacities, and Endogenous Network Formation: the Case of Power Plants and Coal Suppliers in the US*

Click here for the latest version

Yichun Song†

January 14, 2023

Abstract

This paper develops a structural model to study the outcome of a structural change (e.g., merger, entry or exit), which is a critical task in economics. The model incorporates micro-level Nash bargaining, endogenous network formation, suppliers' capacity constraints, and multiple continuous purchase behaviors for a market with business-to-business transactions. I provide a detailed discussion on model identification, implementation, and algorithms for estimation and counterfactual. To operationalize the model, I construct a comprehensive data set of power plants and coal companies from 1998 to 2008. Next, using the data, I estimate the model and evaluate the Arch-Triton merger in 2005, which the FTC initially challenged but failed to block. In contrast with a static network model without cost synergies, my model suggests that the equilibrium prices under a new stable network may increase in the absence of the merger due to capacity constraints. Moreover, divestiture remains a valid merger remedy since equilibrium prices would always increase otherwise.

*I would like to thank Adam Dearing and Matthew Weinberg for their continuous support. Also, I benefited greatly from the discussion with Jason Blevins, Mike Abito, Ohyun Kwon, Hanbat Jeong, Gautam Gowrisankaran and Haopeng Shen. I also would like to thank the seminar participants at of the Ohio State University and Cleveland State University for their helpful comments. All errors are my own.

†Department of Economics, The Ohio State University. Email: song.1399@osu.edu.

1 Introduction and Literature Review

Firms commonly seek to consolidate with others to lower the production cost or obtain new technologies Such consolidation tends to benefit its members but may hurt other parties in the market[1]. Understanding the unintended anti-competitive consequences of mergers is, therefore, a prominent task in industrial organization. The seminal work by Baker and Bresnahan (1985) inspired a broad and rich literature on merger analysis. Nevo (2000), for example, uses the Nash Bertrand competition on the supply side to study the merger effect on posted prices, whereas Houde (2012) extends the framework to allow for spatial competition.

In contrast to the posted price, some important markets are characterized by business-to-business transactions, where the prices are negotiated privately. One example is the market between coal-fired power plants and coal producers. Cicala (2015) shows that plants purchase coal through individual bargaining with suppliers, which induces different behaviors from gas-fired plants after deregulation. However, analysis for structural change in such markets is less developed. This study contributes to the literature by proposing a generalized framework to analyze the consequences of structural changes, such as merger, entry, or exit, when prices are privately negotiated. I extend the current model by allowing a rich set of features, such as endogenous transaction network formation, productive capacity constraints and the coexistence of long-term contracts and spot markets.

This study mainly extends the current merger analysis for individual transactions from two aspects. First, current structural analysis with individual bargaining for industries other than the health care, like, Crawford and Yurukoglu (2012), Grennan (2013), Gowrisankaran et al. (2015) and Gross (2019), are all relied on "static" networks. Crawford and Yurukoglu (2012) studies a short-run welfare effect on TV bundling and shows that distributors' increased cost would offset the welfare gain by offering individual TV channels. Grennan (2013) estimates a bargaining model on a medical device, showing that a largely-neglected bargaining effect is important for the policy outcome of uniform pricing. Gowrisankaran et al. (2015) instead estimate an empirical Nash bargaining model, discuss a merger blocked by the antitrust division and demonstrate that that prices would

1. For example, a merger that reduces competition among upstream suppliers may hurt downstream consumers. As such, the equilibrium prices in the market may even increase as a result of the merger even if it were cost-effective. Moreover, a merger may have a far-reaching implication beyond prices, including product characteristics (Fan, 2013), product proliferation (Fan and Yang, 2020), product promotion (Tenn et al., 2010), advertising strategies (Chandra and Weinberg, 2018). A merger can also trigger tacit price coordination as demonstrated in Miller and Weinberg (2017), Miller et al. (2021).

increase in the presence of merger. Gross (2019) discusses the influence of private labels in the wine industry where wholesale and retail prices are negotiated simultaneously.

However, their models are "static" regarding networks since they do not allow any variations in the parties involved in transactions after a structural change. In contrast, I construct a complete model for buyers' choices, where buyers can choose which sellers to purchase from and the purchase quantities simultaneously[2], which makes the network itself endogenously determined inside the model. The unobserved prices for the non-purchased goods are calculated as the equilibrium solution. Moreover, the sellers that are not currently in the purchase set but still in the choice set of a buyer serve as a threat of replacement, helping the buyer get a better bargaining position. This generates an interconnected purchase pattern inside a market.

The endogenous network formation follows the logic similar to Liebman (2018), Ho and Lee (2019), and Ghili (2022). They expand the Nash bargaining model to allow for an endogenous formation of the transaction networks. Liebman (2018) and Ho and Lee (2019) explore the effect of narrowing the network between hospitals and insurers and show that insurers may deliberately exclude some hospitals to negotiate a better price. The threat of exclusion may render hospitals incapable of charging a higher price. Consequently, a network adequacy law, forcing insurers to spread the coverage of hospitals, would increase the cost. Ghili (2022) goes a step further and allows the threat of replacement to appear also in the estimation rather than just the counterfactual. In Ghili (2022), observed prices are generated by Nash bargaining or the threat of replacement from the currently excluded choices. Furthermore, his counterfactual illustrates that the network adequacy law would decrease the insurers' costs first and then increase when the proportion of hospitals to be included reaches a high level.

For the empirical part of this study, I deviate from the well-researched network expansion in health care and focus on the structural changes in the coal markets in the US. I evaluate the influence of a merger between coal producers retrospectively in 2005. The merger was between a major producer, Arch Coal Inc, and an important fringe producer, Triton Coal Company LLC, of the subbituminous

2. The multiple discrete-continuous choice like Bhat (2005) and Bhat (2008) can accommodate to this also. Train and Wilson (2011) applies Bhat (2005) to model the coal demand of facilities across the US. However, their methods are incompatible with the usual price-endogeneity assumption in industrial organization. The only way the endogeneity can be resolved in their models is through the control function, while the usual control function separability defined in Blundell and Matzkin (2014) prevents the supply side from being Nash bargaining. Kim and Villas-Boas (2018) also points this out and argues that one can still use this by assuming an unrealistic error structure. Thus, I find the following Train and Wilson (2011) may generate inconsistent results.

coal in the Powder River Basin (PRB). This proposed merger was considered anti-competitive and was subject to strict scrutiny by Federal Trade Commission (FTC). FTC attempted but failed to block the merger because the district court accepted the remedy that FTC rejected – the divestiture of Triton's other coal mine, Buckskin, to a company that was previously not in the market for subbituminous coal.

The second major extension I have is tailored but not restricted to this empirical application – producers are capacity-constrained. Presuming that coal companies can supply as much as possible would be unrealistic under environmental regulation and highly-concentrated railroad transportation. Also, it is hard to imagine that plants would refuse coals sold at marginal cost. With capacity constraints, another layer of supplier's optimization is incorporated; sellers also want to maximize their profits conditional on fixed capacities, which, at first sight, renders a different threat-of-replacement from Ho and Lee (2019) and Ghili (2022). As illustrated in the sections below, capacity restrictions would change the market equilibrium and generate distinct counterfactual results. Besides, the algorithms dealing with the capacity constraints, though specially tailored to merger analysis of coal market, may be directly extended and applied to other structural analysis where producers face some restrictions that prevent them from exhausting their production capabilities. Examples include, and not restricted to, raw material bottlenecks (Valero et al., 2018), emergency order opportunity[3] (Zheng et al., 2019), or hold-up problems (Zahur, 2022).

This research contributes to both the applied studies and the methodologies. Firstly, this study adds to the vast collections of merger studies across industries. To my best knowledge, it constitutes the first paper using an endogenous network to study a merger inside a market with private negotiation on price. Also, it is the first study attempting to include all those interactions simultaneously. I demonstrate that a more flexible model indeed makes a difference. For instance, the counterfactual prices could be different by allowing a re-optimized network after a structural change.

Methodologically, though the primary method builds on Ho and Lee (2019) and Ghili (2022), this study has several significant differences. In Ho and Lee (2019), when estimating, they assume that Nash bargaining determines all the observed prices. The threat of replacement only enters into the counterfactual. This framework is limited. Like Ghili (2022), I assume that the observed prices can

3. Actually, in the coal market, the spot market could be viewed as an emergency order opportunity.

be generated either by static Nash bargaining or binding threats that drag down the bargaining prices. Compared with Ghili (2022), firstly, as mentioned above, I introduce the producers' capacity constraints into the model, which is more realistic but requires higher computational complexity. Secondly, I adopt a different strategy for supply-side estimation. Rather than the inequalities and quadratic optimization in Ghili (2022), I turn to a nonlinear equality constraint for the fixed cost of the purchased goods, making the problem a high-dimensional nonlinear optimization. Correspondingly I end up with a different counterfactual algorithm. I develop and solve a larger computational problem than either Ho and Lee (2019) or Ghili (2022), demonstrating the algorithms' practicability in handling complicated problems[4].

The rest of the paper is organized as follows. Section 2 gives an intuitive example, introduces the data for this study and shows the way combine those separate sources as a comprehensive data set. Section 3 contains the Arch-Triton merger's basic timeline and preliminary diff-in-diff regressions about the possible anti-competitive effects. In this case, diff-in-diff gives ambiguous evidence of the price change after the merger, demanding more exploration in structural model. Section 4 examines a structural model for both demand and supply, and Section 5 discusses how the theoretical model becomes empirically estimable and the corresponding computational algorithms. Section 6 shows the estimated results. The following section (Section 7) carries the counterfactual outcomes regarding the prices and facilities' welfare without the merger and without the divestiture. Section 8 is the conclusion. Appendices contain the algorithms, proofs, analysis, and further discussions omitted in the main text.

2 Descriptive Analysis of the model and the Coal Industry

2.1 An Illustration of the Endogenous Network with Capacity Constraint

Here, I illustrate the basic logic behind the model using a parsimonious case with three suppliers and two buyers. Let $P_i(j)$ denote the price of good i sold to buyer j and suppose suppliers only produce one good. In Figure 1, a solid line depicts a realized transaction between buyers and sellers, while a dashed line describes that a buyer does not purchase from a currently active seller. Without loss of generality, one can focus on buyer 2. Buyer 2 purchases from supplier B at the

4. Besides, I stick to point estimation of parameters by imposing an extra selection criteria, which is also different from the partial identified moment inequalities advocated in recent works like Pakes (2010) and Kline et al. (2021). Empirical researches includes Bontemps et al. (2022), which proposes a moment-inequality based two-stage model on the US domestic flights. They evaluate a merger between two airlines and the potential merger remedies.

price $P_B(2)$ and supplier C at $P_C(2)$. If we have Nash bargaining based on a "static" network, then $P_B(2) = P_B^{Nash}(2)$, and as in Grennan (2013), in this case, with $\check{\boldsymbol{P}} = (P_A(1), P_B(2), P_C(2))$[5], $\check{\boldsymbol{Q}} = (Q_A(1), Q_B(2), Q_C(2))$, $P_B^{Nash}(2)$ is the solution to,

$$
\begin{aligned}
&P_B^{Nash}(2) = \underset{P_B(2)}{\arg\max} \left[\pi_B - \pi_{B,-2}\right]^{\mu} \left[\pi_2(\check{\boldsymbol{P}}, \check{\boldsymbol{Q}}) - \pi_2(\check{\boldsymbol{P}}_{-\{2,B\}}, \check{\boldsymbol{Q}}_{-\{2,B\}})\right]^{1-\mu} \\
&\text{with} \quad \pi_B = P_B(2) \times Q_B(2) - TC_2(Q_B(2)),\ \pi_{B,-2} = 0, \\
&\qquad \pi_2(\check{\boldsymbol{P}}, \check{\boldsymbol{Q}}) = TS_2(Q_B(2), Q_C(2)) - [P_B(2) \times Q_B(2) + P_C(2) \times Q_C(2)], \\
&\qquad TS_2(\check{\boldsymbol{P}}_{-\{2,B\}}, \check{\boldsymbol{Q}}_{-\{2,B\}}) = \max_{Q'_C(2)} TS_2(Q_C(2)) - P_C(2) \times Q'_C(2), \\
&\qquad \mu \text{ is the bargaining leverage, } TC \text{ is the total cost, } TS \text{ is the total surplus.}
\end{aligned}
\tag{1}
$$

A static Nash bargaining framework treat the prices as the solution to the problem in (1). Specifically, the disagreement payoff for buyer 2 is the surplus after a failure in the bargaining with supplier B and a re-optimizing over the purchase quantity at current price $P_C(2)$. In contrast, in this study, firstly I allow for an endogenous network formation, where buyer 2 can use the currently non-purchased supplier A's good as a threat when bargaining with supplier B. However, instead of incorporating this threat in buyer 2's disagreement payoff [6], I treat it as a restriction in the maximum price that B can charge, or, $P_B(2) = \min\{P_B^{Nash}(2), P_B^{NR}(2)\}$. NR is the abbreviation for "no-replacement" and $P_B^{NR}(2)$ is the maximum price supplier B can charge to buyer 2 without inducing a higher surplus when buyer 2 switch to supplier A instead. Intuitively, $P_B^{NR}(2)$ would depend on $P_A(2)$. For instance, without capacity constraint, in the short run, supplier A would want to sell to buyer 2 at any price larger than its marginal cost $MC_A(2)$.

While the endogenous network formation bounds the prices by $P_B(2) = \min\{P_B^{Nash}(2), P_B^{NR}(2)\}$, supply-side capacity constraints, or the maximum quantities, determine where the $P_B^{NR}(2)$ lies[7]. In the right panel of Figure 1, suppliers face upper limits $\bar{Q}_A$, $\bar{Q}_B$, and $\bar{Q}_C$, restricting the quantities they can produce. Correspondingly, a network with both endogenous network and capacity constraint would have $P_B(2) = \min\{P_B^{Nash}(2), P_B^{NR}(2)\}$, and $P_B^{NR}(2)$ is jointly determined by $P_A(1)$, $P_A(2)$, $\bar{Q}_A(1)$ and $\bar{Q}_A(2)$. $\bar{Q}_A(2)$ is the corresponding sales quantity when evaluated at

5. The prices associated with dashed lines, like $P_A(2)$ and $P_B(1)$, are irrelevant in a static bargaining game.

6. Putting the threat in the disagreement payoff of equation (1) and replace the $\pi_2(\check{\boldsymbol{P}}_{-\{2,B\}}, \check{\boldsymbol{Q}}_{-\{2,B\}})$ is possible. I decide not to proceed on with this due to a computational reason after fixed costs of inclusion a currently not-purchased good are introduced.

7. Capacity constraint is one of many other factors that prevents the suppliers from selling as much quantities as they want. Thus, the method developed in this study can be directly generalized into other scenarios.

current situation. The minimum acceptable price for supplier A, or $P_A(2)$, equals $MC_A(2)$ whenever at $MC_A(2)$, $Q_A(1) + Q_A(2) \leqslant \bar{Q}_A$. Otherwise, $P_A(2) > MC_A(2)$, since to sell to buyer 2, A must cut down some current sales to buyer 1, and usually in an oligopolistic market, there is $P_A(1) > MC_A(1)$.

Consequently, with both endogenous network formation and suppliers' capacity constraints, the transaction pattern becomes an intriguing interconnected network with realized prices possibly affected by every component inside. That is the reason I consider this as an "endogenous network" in this study rather than a bilateral contracting game.

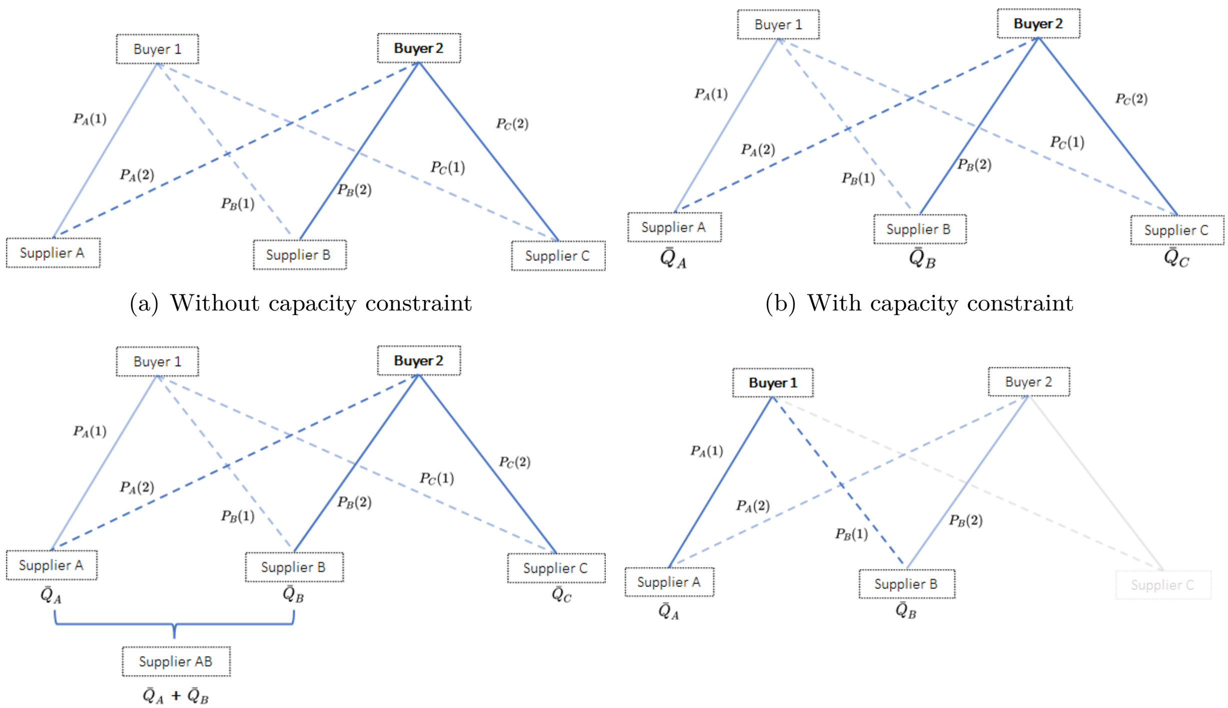
Moreover, this flexible network structure would generate a different counterfactual simulation whenever a structural change happens. Suppose there is a merger between supplier A and B, as indicated in Figure 1, part (c), after which, supplier A and B operate together and has $\bar{Q}_A + \bar{Q}_B$ as the new capacity constraint. In a static Nash bargaining in (1), this merger would not influence the resulted $P_B(2)$ since good A is not in the purchase set of buyer 2, so is not part of the disagreement payoff. However, with network formation, this may influence $P_B(2)$ through $P_B^{NR}(2)$, where it would be weird for buyer 2 to threat to switch to supplier A whenever the $P_B^{Nash}(2)$ is too high. If $P_B(2) = P_B^{NR}(2)$ and $P_B^{NR}(2)$ is determined by $P_A(2)$[8], then a merger between A and B would increase the price paid by buyer 2 [9]. This may lead to chain reactions since, for instance, now the disagreement payoff for buyer 2 when bargaining with supplier C would decrease with a potentially increased $P_B(2)$.

Finally, Panel (d) in Figure 1 contains the case where supplier C drops out of the market – a situation that static framework cannot handle. Here, consider the buyer 1. Similarly, the exit of supplier C causes the loss of a possible threat of buyer 1, which would increase $P_A(1)$ if $P_C(1)$ constitutes a binding threat. And supplier B, with a fixed $\bar{Q}_B$ may find it hard to satisfy the quantity demanded by buyer 2. Consequently, $P_B(2)$ may also increase and in the short run, there could be shortage in the market.

2.2 Data Description

I constructed a comprehensive data set containing both information from the demand and supply sides. The data set is composed from a variety of different sources. The monthly fuel transaction

8. which is the case here since good A is the only choice that buyer 2 does not purchase from.
9. If, before the merger, $P_B(2) = P_B^{Nash}(2)$, then the disappearance of threat from A does not matter.



(a) Without capacity constraint

(b) With capacity constraint

(c) With a potential merger

(d) With supplier C exit

Figure 1: An illustration of a simple network.

data comes from form EIA 423, FERC 423 and EIA923, collected and organized by S&P Platts[10]. EIA and FERC began to collect monthly fuel transaction data of power plants with a fossil-fueled nameplate generating capacity $\geqslant$ 50 MW [11] since the 1970s. The data I use here spans around ten years, from Jan 1998 to March 2008. A transaction record carries some detailed information, for instance,

"George Neal South, a power plant belonging to MidAmerican Energy Co, on March 2008, purchased SPOT contracts from the Black Thunder mine in the South PRB region owned by Arch Coal, a total of 95340 short tons, containing 8863 Btu/Lb, 4.8% of Ash. 0.54 Lb of SO_2 would be generated every MMBtu. And it paid \$20.17 per short ton, including the \$16.8 per ton shipping cost, and the Free-on-Board price (abbreviated as FOB price)[12], for the data point above, is around \$3.37 per short ton. "

10. Several differences between the public-available EIA/FERC data and the public-available S&P Platts data: (I) Platts has transaction data for both regulated and deregulated plants, where only the cost-regulated plants are ready in the publicly available data set on the EIA web page; (II) Platts has the overall prices decomposed into the Free-on-Board price and the shipping cost, which does not exist in the EIA original data; (III) while the EIA/FERC data reports every transaction that happened, Platts has already aggregated them up into monthly level.

11. EIA webpage for 423: https://www.eia.gov/electricity/data/eia423/

12. The prices charged to the power plant at the mine-mouth.

Besides the transaction-specific data, characteristics and operation status inside a plant would affect its purchase behavior. First, EIA Form 767, 860 and 923 contain information about the installation of scrubbers[13], an equipment among the most crucial capital investments a power plant can install to comply with environmental regulations. For a generating unit, scrubber installation is treated a binary outcome, where I abstract away from the differences among products. As for plants' revenues and other operation costs, I extract the state-level monthly average revenue from selling electricity from EIA-861M/862 [14]. Furthermore, to model the surplus-optimizing behavior of power plants, rather than minimization of the fuel procurement cost, a rough proportion of fuel cost towards facilities' operating cost is needed. For this, I refer to the FERC Form 1, a publicly available data[15]. For the plants that do not respond to Form 1, I have EIA form-412 [16] serving as a complement to Form 1. I then impute the data for plants that do not appear in either records[17].

Fuels purchased by power facilities serve as an intermediate materials for electricity generation. The quantities of fuels needed lies on the requirement of power generation. CEMS high-frequency generation data collects the emission of wastes from the power plants that participate in at least one of those Air Market Programs[18]. The data relevant to this study is the high-frequency hourly gross electricity generation for each coal-fueled generating unit in the power plants. Additionally, each power plant belongs to a specific NERC region[19]. Distinct characteristics inside a NERC

13. EIA Form 767 contains generation configuration data before and for 2005; after a gap in 2006, EIA 860 and EIA 923 took over the job. The historical data in EIA 767 is in https://www.eia.gov/electricity/data/eia767 and EIA 860 in https://www.eia.gov/electricity/data/eia860. Moreover, I perform a cross-check using the arranged data set in Cicala (2015).

14. Source: https://www.eia.gov/electricity/data/eia861m/#salesrevenue. Unfortunately, the revenue specific to power plant is unavailable for many plants. The state-level quantity-averaged revenue is the best approximation I can find.

15. Form 1 can be downloaded and viewed in its original electronic form through the software provided officially, from https://www.ferc.gov/industries-data/electric/general-information/electric-industry-forms/form-1-electric-utility-annual.

16. Available at https://www.eia.gov/electricity/data/eia412/, provides annual financial reports, with usually detailed cost information included, until 2003. EIA has terminated its 412 data collection since 2003.

17. Although files sent towards US Securities and Exchange Commission (SEC) are publicly available, one problem for using these SEC 10-K or 10-Q is that they do not provide separate financial information for the coal-fired power plants. This lack of plant-level data could generate an imprecise estimation of the proportion of fuel cost for corporations owning nuclear or hydroelectric stations simultaneously. Therefore, I stick to the information summarized from Form 1 and EIA 412.

18. At the Air Markets Program Data website, https://ampd.epa.gov/ampd/. The Air Market Programs are the programs involving environmental protection; includes, for example, Acid Rain Program (ARP), Mercury and Air Toxics Standards (MATS), Clean Air Interstate Rule SO2 (CAIRSO2), and so on. Usually, moderate to large power plants would have a high probability of participating in some environmental programs.

19. NERC is the abbreviation of North American Electric Reliability Corporation, designed to ensure the reliability of the bulk electric power system in North America. Inside a NERC region, planners should balance their electricity demand and supply. Governors would anticipate almost no electricity transaction across the boundary, making each NERC region a self-supported area.

region may influence plants' fuel decisions. The division of the NERC region evolves over the years, with data provided in the eGrid[20].

Some factors outside a coal-fired plant may also influence the coal consumption. Natural-gas-fired plants are significant rivals for the coal-fired plants during the sample period, though coal-fired plants still took a dominant position, as shown in Figure 2. I obtained daily natural gas price of many hubs across the US from Natural Gas Intelligence[21]. The high-frequency daily natural gas data, combined with the high-frequency electricity generation data, helps to evaluate a short-run reaction towards relative fuel costs. Also, the Acid Rain Program (ARP) has introduced an allowance trading system to reduce outflow of sulfur dioxide (SO_2) of power plants. Consequently, power plants would be sensitive to coal's sulfur content adjusted by the current prevailing sulfur dioxide allowance in the market.

There is no unique identifier for power plants across all data sets. Hence, I link the data by manually generating a reference code between the Plant ID in Platt's transaction data and the ORISPL identifier in the rest of the data sets, using the plants' names, geographic locations, and operator company.

2.3 Background

Electricity in US is generated with a variety of fuels. According to the EIA, "...three major categories of energy for electricity generation are fossil fuels (coal, natural gas, and petroleum), nuclear energy, and renewable energy sources. Most electricity is generated with steam turbines using fossil fuels, nuclear, biomass, geothermal, and solar thermal energy." Figure 2 shows the major sources of power generation in the US from 1995 to 2014, during which coal-fired plants hold an outstanding share of electricity generation. However, there is a declining trend since 2006 – the natural gas plants gradually took over the proportion of coal-fired electricity. This downward trend persists till now and in 2020, the share of coal dropped to 19.3%. Nevertheless, during the period of the data I have, the coal-fired power plants still occupied a predominant place.

Unlike natural gas, often viewed as a homogeneous commodity (Davis and Kilian (2011)), coals are differentiated. There are four types of coals globally, three of which are prevalent in the US,

20. eGrid data source: https://www.epa.gov/egrid/download-data, for intermittent years of 1998, 1999, 2000, 2004, 2005, 2007.

21. Daily Historical data in Natural Gas Intelligence webpage: https://www.naturalgasintel.com/product/daily-historical-data/.

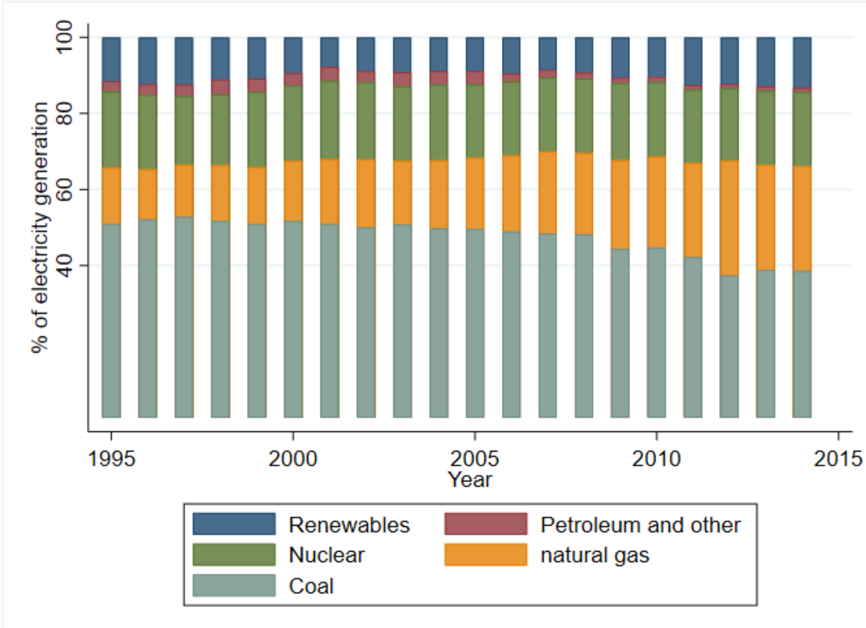


Figure 2: The proportion of each source for the electricity generation. Source: EIA Website, https://www.eia.gov/energyexplained/electricity/electricity-in-the-us.php

as summarized in Table 1. In summary, the differences between coals come from three aspects: energy content[22], and two impurities – sulfur and ash content. Most coals burned in the US are either bituminous coal – moderate to high content in energy, sulfur, and ash, or subbituminous coal – low to moderate energy content and relatively lower pollutant levels.

The use of subbituminous coal has been growing in the electricity sector after the phase-in of Clean Air restriction. Although it requires advanced techniques for efficient combustion, subbituminous coal generates less sulfur dioxide, beneficial to plants that have a problem installing the capital-intensive scrubbers or a hard time in purchasing quota to support their sulfur dioxide emission. Bituminous coal and subbituminuous are produced in different areas. Bituminous coal comes from mines all over the nation[23], while the majority of subbituminous coals originates from the Powder River Basin (PRB), located in Wyoming (Southern Powder River Basin, or SPRB) and Montana (Northern Powder River Basin, or NPRB).

Table 10 summarizes Platt's transaction data by major coal supply regions. Consistent with Table 1, coals from PRB have lower carbon content but are more environmental-friendly. NPRB and

22. Or roughly the carbon content, the active ingredient in coal where the energy comes.
23. Major production areas of bituminous coal include: Central Appalachia, Northern Appalachia, Mid-Continent Illinois Basin, Western Central Rockies, and Western Four Corners.

	Bituminous	Subbituminous	Lignite	Anthracite
Fixed Carbon Content	45% to 86%	35% to 45%	25% to 35%	86% to 97%
Sulfur Content	0.7% to 4.0%	< 2.0%	0.4% to 1.0%	0.6% to 0.8%
Ash Content	3% to 12%	< 10%	10% to 50%	10% to 20%
Proportion of total sales	Largest by energy content	Largest by total quantity	Less than 10%	About 0.2%
Locations in the US	All over US, Appalachia, Western	Wyoming Montana	Texas North Dakota	All in Northeast Pennsylvania
Mainly usage in the US	Electricity, Steel	Electricity	Electricity	-

Table 1: Four types of coal, and how they are differentiated with each other. Source: EIA Website, https://www.eia.gov/energyexplained/coal/ and https://www.purdue.edu/discoverypark/energy/assets/pdfs/cctr/outreach/Basics8-CoalCharacteristics-Oct08.pdf.

SPRB coals have comparable characteristics. PRB coal is cheaper than non-PRB coal, even after accounting for the differences in the energy content[24]. Table 10 also indicates that some of the gaps between FOB price are wiped away after the transportation expenses added up. With non-PRB coal spreading all over the US, plants can always choose the closest non-PRB mine to save the transportation cost. However, for PRB coal, power plants have to accept the long transportation distance and a higher average shipping cost.

Even when the transportation cost account for a significant fraction of the total cost, facilities all over the US purchased PRB coal from 1998 to 2008. Figure 3 shows the proportion of PRB-purchasing plants. A lighter color means a small proportion of plants operating in that state using PRB coal. PRB coal would be the choice for almost every coal-fired plant in states near the production region. The states with a low PRB adoption would either be far from the production area or close to some other supply region of non-PRB coals.

From a descriptive view, the market structure for PRB suppliers is different from non-PRB suppliers. The PRB coal is dominated by a few big companies, while the market for non-PRB coal is more competitive. Figure 4 shows the major suppliers in PRB and non-PRB regions during the

24. One possible reason for this dispersion is the different marginal cost of production. PRB coal is buried shallower under the ground, which can be handled by a cheaper and safer surface mining technique. On the contrary, underground mining may be inevitable for many non-PRB mines. This difference in marginal cost drives up the gap between the FOB prices

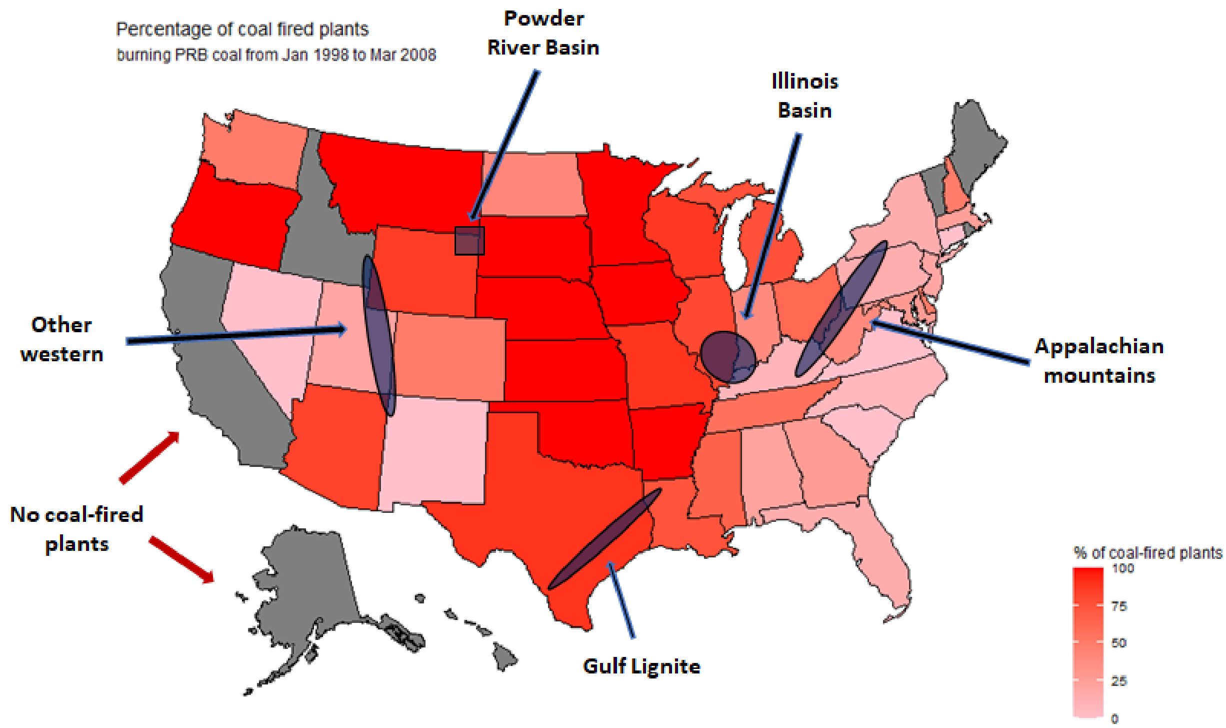


Figure 3: State-level proportion of coal-fired power plants purchasing PRB coal, from 1998 to 2008.

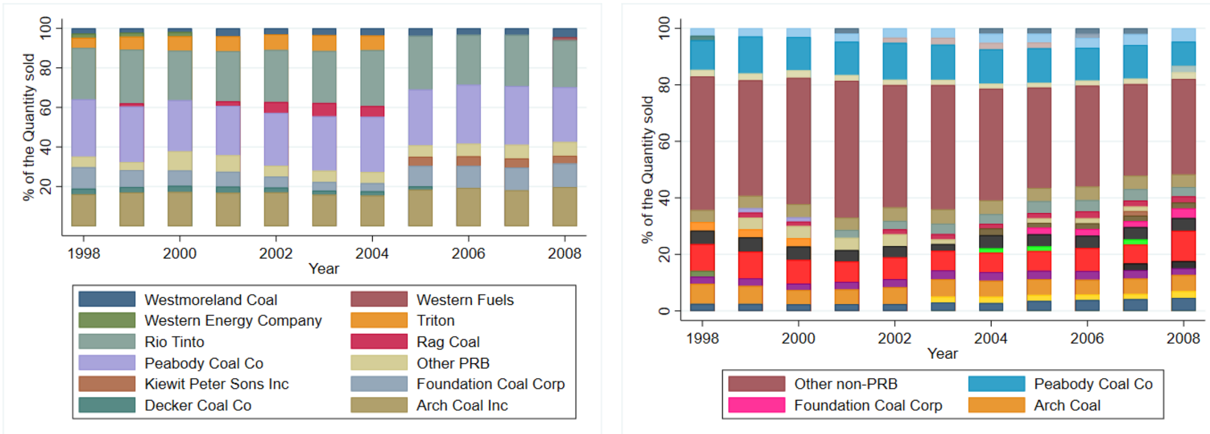
sample periods. In the PRB market, Peabody Coal Co accounts for more than 20% of the total sales. Arch Coal Inc, the dominating participant in the merger with Triton, occupies a market share slightly below 20%. Another vital player in PRB is Rio Tinto, whose share in PRB coal is comparable to Peabody. The aggregation of all other small suppliers in PRB held around 5% of sales. In contrast, despite a non-negligible proportion of Peabody and Arch, the Other non-PRB occupies a large share, especially during the first half of the period, leading non-PRB coal to be a competitive market. Thus, a proposed merger in PRB coal would naturally attract more attention.

3 Arch-Triton Merger

3.1 Time line of Arch-Triton Merger

On March, 2003, Arch Coal, Inc (Arch) entered into a merger agreement with Triton Coal Company, LLC (Triton), owned by New Vulcan Holdings, LLC. While at that time, the PRB coal market was dominated by the Big Three – Arch, Peabody, and Rio Tinto[25], Triton, holding two coal mines,

25. One can find a summary of Arch-Triton merger in Commentary on the Horizontal Merger Guidelines, 2006. https://www.justice.gov/sites/default/files/atr/legacy/2006/04/27/215247.pdf.



(a) The major coal companies for PRB supply region (b) The major coal companies for non-PRB supply region

Figure 4: Difference in Market Structure, of PRB coal and non-PRB coal.

North Rochelle and Buckskin, was an important fringe producer. In 2003, North Rochelle and Buckskin were the No.7 and No.10 largest coal mine in the US [26]. Arch, owning the Black Thunder, the second-largest coal mine in the US, was a significant producer. In 2003, the total quantities produced by Black Thunder were about 1.5 times of the summation of Triton's production.

The primary goal for which Arch raised this merger is cost-saving[27]. Geographically, North Rochelle and Black Thunder are adjacent to each other. Arch claimed that this integration of North Rochelle and Black Thunder would create tremendous cost-savings through synergies. However, on April, 2004, FTC filed a complaint to block this proposed consolidation. FTC describes Arch-Triton merger as follows[28],

"The Commission authorized staff to file a complaint to block Arch Coal, Inc.'s proposed acquisition of Triton Coal Company, LLC from New Vulcan Holdings, LLC on grounds that the acquisition would increase concentration and tend to create a monopoly in the market for coal mined from the Southern Powder River Basin and in the production of 8800 British Thermal Unit coal. On April 1, 2004, the complaint was filed in the US District Court for the District of Columbia; the court denied the FTC's motion for a preliminary injunction. On June 13, 2005 the Commission announced that it was closing its investigation, saying that it will not continue with administrative

26. And also in the PRB, because the ten biggest coal mines in the US are all PRB mines.
27. This is indicated by Steven F. Leer, Arch Coal's president and chief executive officer. The news & media page for this merger can be found in https://investor.archrsc.com/news-releases/news-release-details/arch-coal-completes-acquisition-triton-coal-company.
28. The webpage for the Arch-Triton merger, in FTC website, https://www.ftc.gov/enforcement/cases-proceedings/031-0191/arch-coal-inc-new-vulcan-coal-holdings-llc-triton-coal

litigation challenging the deal."

Finally, FTC's attempts to stop this transaction worked out in vain, since its complaint was turned down by the US District Court of the District of Columbia. The court accepted a remedy denied by FTC – the divestiture of Buckskin, the smaller one among the two Triton's mines, to a previously marginal producer in PRB, the Kiewit Corporation (Kiewit). Divestiture is an essential solution for mergers between multi-product companies with a possible sizeable anti-competitive outcome. Successful cases include Friberg and Romahn (2015)[29] and Tenn and Yun (2011). On the other hand, divestiture could disappoint the Anheuser-Busch and Grupo Modelo Merger in the US beer industry due to case-specific reasons like binding capacity constraints (Wang et al., 2017). In this Arch-Triton merger, despite this possible divestiture remedy, FTC still thought that transforming the productive capacity from an important fringe firm to Big Threes would significantly weaken the competition, and even create a monopoly in the PRB region.

The merger happened when coal-fired facilities still absorbed the largest fraction of power demand in the US. This Arch-Triton merger also raised concerns from their industrial consumers, who rely on PRB coals as the primary fuel in their generating units[30]. Hence, the debatable nature of this merger makes it an ideal case for an application of a more realistic structural setting.

3.2 Descriptive Figures

As shown in Figure 5, after FTC's close of investigation, the real FOB price for PRB coal indeed increased. In comparison with a downward sloping trend in the non-PRB suppliers, the median spot prices of PRB transactions displayed a sudden jump. Prices in the long-term contract increased gradually following a lagged pace of spot market.

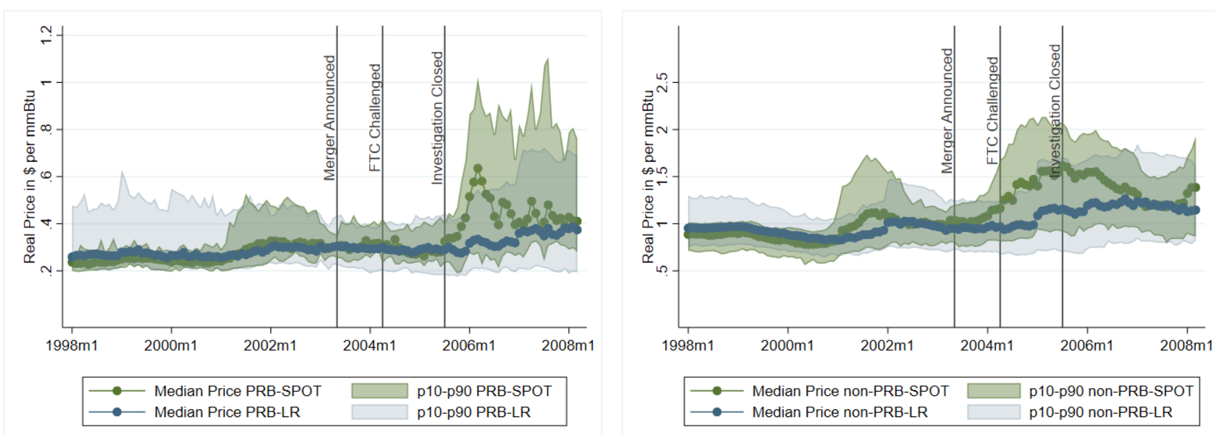
Though, the interpretation of this price increase can be tricky. Firstly, the trend of median prices deviated between non-PRB and PRB coal from the time the merger was proposed, so that there could be some strategic behaviors among entities engaging in this merger. PRB coal companies might strategically restrain from price increasing to leave a good expression to the antitrust division. Therefore, the sudden FOB price increase after investigation closed may be a compensation to the previous suppression. Besides, a train derailment happened in May 2005, immediately after the

29. If the divestiture is evaluated using a basic structural model, like a Bertrand competition supply side, then the direction of the after-divestiture price change is fixed. Divestiture would always improve the level of competition. The main interest lies in how large this effect is.

30. Examples like Doyle Jr. and Ray (2005)

merger[31], causing temporary halt in the only railroad leading out of PRB. Coal supplies in PRB did not fully recover before April 2006. This difficulty in coal shipping would be a disturbing factor for the isolation of merger effects. Despite this, the real FOB price was still in a relatively high position, compared with its non-PRB counterpart, since the year 2007.

Moreover, compared with non-PRB coal, Figure 5 shows that the FOB price dispersion in PRB coal also increased after the close of investigation. This leads to possible heterogeneity in price determination. Figure 6 shows that plants that did not purchase from the merged entities, or not from both of the merged entities, still suffer from an increased price. This pattern of price increase across all power plants, regardless of their purchase histories, indicates the possibility of spill-over effect. The price increases for plants that do not have both Arch and Triton in the purchase set cannot be explained if I only have a static network for Nash bargaining. In a static Nash bargaining setting, power plants that did not purchase both face no structural change in their threat points after the merger. Thus, Figure 6 indeed provides a descriptive justification for the need of endogenous network formation.



(a) PRB coals

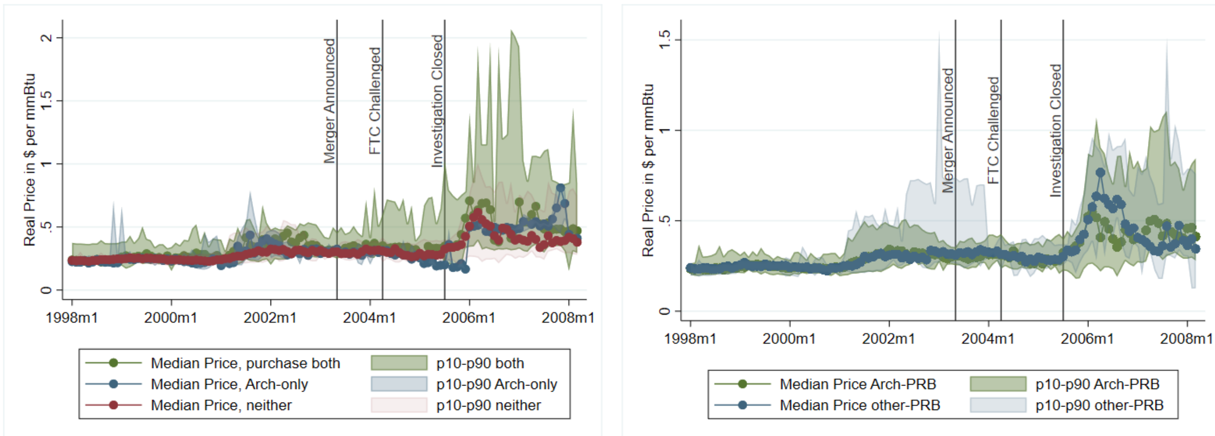
(b) non-PRB coals

Figure 5: Price change after Arch-Triton merger.

3.3 Reduced Form Difference-in-Difference Estimation

Then I carry on difference-in-difference (diff-in-diff) regressions to further explore the prices change after the merger. The ambiguous results from this section calls for the need of structural analysis.

31. Deliveries of Coal from the Powder River Basin: Events and Trends 2005-2007, https://www.oe.netl.doe.gov/docs/Final-Coal-Study_101507.pdf



(a) Price change for different types of plants (b) Price change for Arch and non-Arch coals

Figure 6: Price change after Arch-Triton merger, for different types of power plants. Types are generated from their purchase histories.

The basic structure takes the form,

$$
\begin{aligned}
p_{ijt} = \theta_i + \tau X_{ijt} + Year_t + Month_t + \gamma_1 \times Post_t + \gamma_2 \times Post_t \times PRB_j \\
+ \gamma_3 \times Post_t \times Tier1\, PRB_j + \gamma_4 \times \mathbf{1}(Spot) \times Post_t \times PRB_j + \epsilon_{ijt}
\end{aligned}
\tag{2}
$$

P_{ijt} can be real level price or the log price, or quantities-weighted unweighted prices. X_{ijt} are other controls that possibly change the prices and the baseline fixed effects for PRB coal, Tier 1 PRB coal, etc. θ_i, $Year_t$ and $Month_t$ are dummy variables for plants, year, and month. We are interested in the coefficients for the interaction terms, γ_2 to γ_4. Both Black Thunder and North Rochelle belong to the subclass called Tier 1 PRB coal [32], so I may expect Tier 1 PRB coal would have different price trends.

To account for the unobserved heterogeneities, I use the a matched diff-in-diff setting as in Heckman et al. (1997) and Dehejia and Wahba (1999), where I compare the SPRB-fired plants with its non-SPRB-fired neighbors. Plants are matched by geographical locations. I match the plants with its nearest one, three, or five neighbors. Figure 7 shows the distribution of distances with different numbers of neighbors enters in. The matched diff-in-diff is conducted via weighted least squares,

32. One may divide the SPRB coal into three sub-classes, namely Tier 1, Tier2, and Tier 3, according to, for example, information provided in https://www.ftc.gov/news-events/news/press-releases/2004/04/ftc-files-federal-complaint-challenging-arch-coals-proposed-acquisition-triton-coal-company. Tier 1 is considered as highest quality. Tier 1 coal includes mines like Black Thunder of Arch, North Rochelle of Triton, Jacobs Ranch and Antelope of Rio Tinto, and North Antelope Rochelle of Peabody. Tier 2 coal distributed to the north of Tier 1, includes Coal creek of Arch, Caballo, Belle Ayr and so on. Tier 3 locates the north of Tier 2. Buckskin of Triton belongs to Tier 3.

with weights w_i generated as

$$w_i = \begin{cases} 1, & \text{If i is SPRB plant} \\ 0, & \text{If i non-PRB plant never matched} \\ \frac{1}{m} \times \left[\# \text{ of SPRB plant it matches}\right], & \text{if i is a non-PRB plant matched} \end{cases}$$

Table 2 and 3 contain the results from the matched diff-in-diff with three nearest neighbors and variance clustered at plant level. Now the dependent variable p_{ijt} is the quantity-weighted values across different goods. Similar to Cicala (2015), deregulated power plants tend to have lower purchase prices. In Table 2, a positive and significant γ_3 shows a price increase for Tier 1 PRB coal after the merger of Arch-Triton. In contrast, other coals from SPRB experienced insignificant or small changes. On the other hand, the signs of coefficients estimated flip when the level price is used as a dependent variable rather than the log prices. This indicates possible undetermined effects of this merger towards price changes.

However, this result may not be robust. First, matching itself could generate biased estimation. To correct this, I also provide results from non-matched samples. The results turn out to have little difference. Other factors may also effect the prices. For instance, in May of 2005, a train derailment occurred on the only track leading out of the Southern PRB, which halted coal shipping until March 2006. This derailment constitutes another possible explanation for this relative price change in PRB coal purchase. Also, a sudden increase in demand for coals produced near the East Coast from foreign countries occurred in December 2007. Results from these diff-in-diff regressions where I remove the data points from those two suspicious periods can be found in Appendix C.1, which does not offer significantly-different results.

4 The Endogenous Network with Capacity Constraint

To further analyze the influence of this Arch-Triton merger and policy implications, I establish a comprehensive structural model fully grasping the transaction relation between consumers and suppliers. This model allows, firstly, power plants making both to purchase or not, and how much to purchase decisions among differentiated goods simultaneously. The observed prices are determined either by Nash bargaining or the threat of replacement. Besides, supply-side capacity constraints determine the price imputation of the non-purchased goods in the equilibrium.

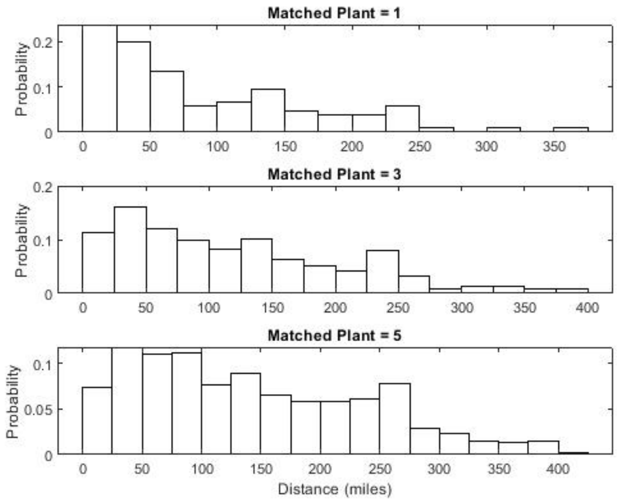


Figure 7: Distribution of distance when matched with different number of neighbors.

	Matched		Not Matched	
Deregulation	-0.3178***	-0.2659**	-0.1054***	-0.0173
	(0.1037)	(0.1040)	(0.0405)	(0.0409)
γ_1	-0.0121	0.1523***	-0.0401***	0.1605***
	(0.0449)	(0.0499)	(0.0099)	(0.0132)
γ_2	-0.0326	-0.0312	-0.0386**	-0.0427**
	(0.0286)	(0.0292)	(0.0192)	(0.0198)
γ_3	0.1375***	0.1339***	0.1354***	0.1330***
	(0.0369)	(0.0370)	(0.0354)	(0.0361)
γ_4	0.0306	0.0354	0.1684***	0.1787***
	(0.0870)	(0.0854)	(0.0616)	(0.0633)
Year-FE	Yes	No	Yes	No
Month-FE	Yes	No	Yes	No
R2	0.8936	0.8901	0.8078	0.7964

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2: Matched Diff-in-diff, in log price.

	Matched		Not Matched	
Deregulation	-0.1762***	-0.1360**	-0.0515*	0.0386
	(0.0575)	(0.0535)	(0.0300)	(0.0299)
γ_1	0.1017*	0.2443***	-0.0124	0.2049***
	(0.0559)	(0.0707)	(0.0096)	(0.0163)
γ_2	0.0669***	0.0652***	0.0663***	0.0648***
	(0.0172)	(0.0173)	(0.0148)	(0.0154)
γ_3	-0.1060***	-0.1069***	-0.0852***	-0.0907***
	(0.0314)	(0.0324)	(0.0102)	(0.0107)
γ_4	-0.0604	-0.0583	0.0267	0.0357
	(0.0687)	(0.0685)	(0.0380)	(0.0395)
Year-FE	Yes	No	Yes	No
Month-FE	Yes	No	Yes	No
R2	0.8997	0.7964	0.9361	0.8689

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3: Matched Diff-in-diff, in real level price.

This section focuses on developing the theoretical model in the supply side. In the supply model, I assume complete information such that ξ_{ijt} is known to both entities. The threat of replacement must be credible since I presume that suppliers can distinguish between the credible and incredible threats instantaneously. Thereby, any threat proposed by the power plant incompatible with the incentive would be discarded. More details would be in the rest part of this subsection.

Here, coal procurement is through individual transaction. And a market is defined as a plant i - time t combination, with participants as coal companies m and power plant i.

Network. A network T is composed by a set of nodes and links, with links connecting nodes. Nodes stand for the participating entities in the market and links describing the relationship among them. Links d_{ijt} have two statuses, solid or dashed. A d_{ijt} is solid if $d_{ijt} = 1$, meaning buyer i and supplier j has built up a temporary transaction partnership at time t. Dashed links refer to the case $d_{ijt} = 0$, where buyers and sellers are still in the market, yet the transaction does not happen between them.

One can find a sample network in Figure 8. Power plants are in the demand side, each belonging to an operating company. In this case, to make a tractable network, the supply-side contains major coal sources from the PRB region and two composite goods – one for Other PRB and one for Other

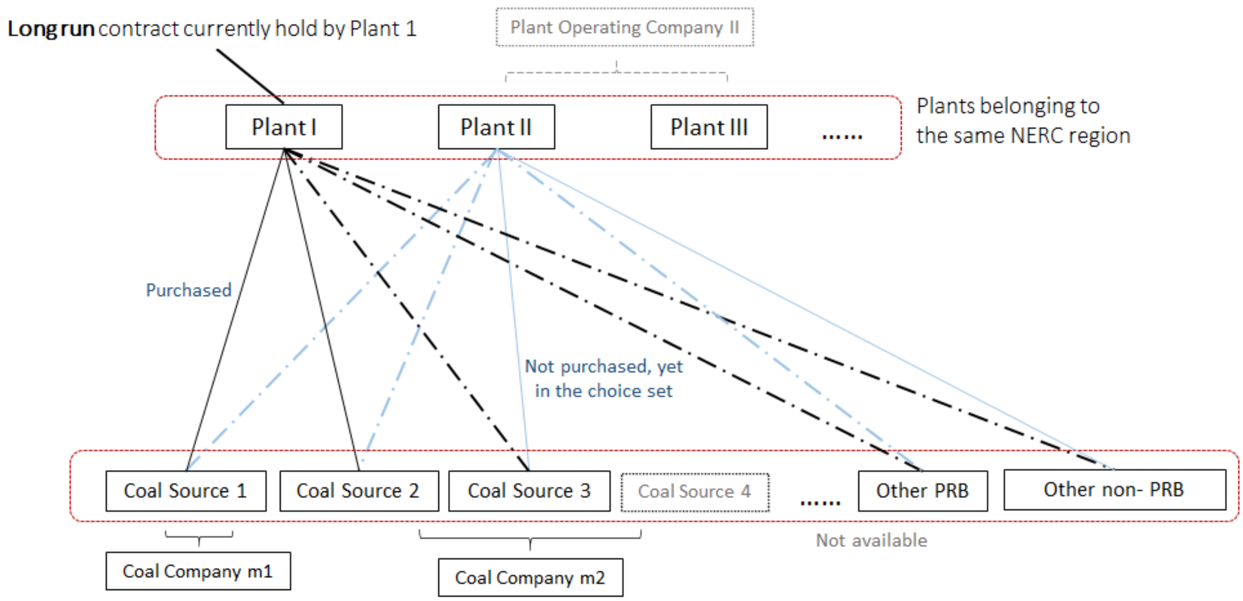


Figure 8: An illustration of a possible network inside a market, where the market is defined by a separate NERC region.

non-PRB. Inside this sample network, Coal Source 4 does not enter into the market at the current period, for some reasons not observed by researchers. Plant II purchases from Coal Source 3, but Plant I does not, so there is a dashed line connecting Plant I and Coal Source 3. Power plants may hold long-term contracts while still entering into the spot market. Currently, the dynamic process to determine long-term contracts is beyond the model's scope, and I treat the them as an exogenous process.

A network structure T implies a set of equilibrium prices P. For a given network T, P is defined as the set of mutually best-response prices. With all others' prices set still, no buyer-supplier pair would want to deviate from it. The deviation here has a two-fold meaning – no one wants to charge a different price for an established link, and no one wants to unilaterally have a different network structure. Certainly, not all randomly generated T could be matched with a set of equilibrium prices, but a stable network would do.

Stable Network. Similar as the overwhelming majority work in economics, what I am concerned about is a network that behaves in an equilibrium way, described by stability. A stable network must satisfy several equilibrium conditions. Succinctly, here a stable network is robust to all possible one-slot (addition or cut-off) and two-slot deviations (replacement) acceptable to both power plants and coal companies. Consistent with many researchers in the structural model, I assume the realized purchase pattern a stable network, which constitutes the foundation for the supply-side estimation. In addition, a stable network is the place the counterfactual analysis arrives at finally when a structure change happens.

One drawback of individual purchase data with differentiated goods is that it is often partially observed. Only the actually-happened transactions are recorded and the prices at which the transaction failed need to be imputed. An imputation that is consistent with the equilibrium is the reservation price p_{ijt}^{RES} calculated inside the equilibrium network. For a network (T, P), the reservation price of coal company m is the minimum acceptable price for selling j to i at t, or the reservation price p_{ijt}^{RES}.

Conditions for a stable network. There are three conditions characterizing a stable network, which restrict any possible unilateral deviation.

Condition I: No addition. Suppose good j' is available at time t, but not purchased by power plant i, then the no addition conditions indicate that power plant i would not want to purchase j', even at the minimum price acceptable to j' s producer. Mathematically, the plant i's overall surplus, contingent on the current network (T^*, P^*), is denoted by π_{it}. Then there must be,

$$\pi_{it}(T^*, P^*_{\{\forall\, p_{ijt} \geqslant p_{ijt}^{RES}\}}) \geqslant \pi_{it}(T^*_{+ij't}, P^*_{+p_{ij't}^{RES}}) \tag{3}$$

$T^*_{+ij'}$ refers to network T^* with the addition j' by i. $P^*_{+p_{ij't}^{RES}}$ denotes that the inserted price for the added good j' is the reservation price of the producer, $p_{ij't}^{RES}$. Namely, there is no price $p_{ij't}$ larger than or equal to $p_{ij't}^{RES}$ that could make the addition of j' profitable to i at t. Also, an implicit condition exists in (T^*, P^*) such that all the prices p_{ijt} in the original network P^* need to be larger than their reservation prices.

Condition II: No Cut-off. Suppose good j is currently in the purchase set of i at time t, then the no cut-off condition represents that plant i would like to cut the good j off from the choice set. Alternatively, cutting off a currently existing link j would not benefit any plant.

$$\pi_{it}(T^*, P^*_{\{\forall\, p_{ijt} \geqslant p_{ijt}^{RES}\}}) \geqslant \pi_{it}(T^*_{-ijt}, P^*_{-ijt}) \tag{4}$$

where T^*_{-ijt} is the network without the solid link ijt, and P^*_{-ijt} is the recalculated equilibrium price after the removal of link ijt.

Condition III: No replacement. Suppose good j is currently in the purchase set of i, j' is not. No replacement means plant i would not replace j with j', at any price at least as large as the reservation level $p_{ij't}^{RES}$.

$$\pi_{it}(T^{*}, P^{*}_{\{\forall\, p_{ijt} \geqslant p_{ijt}^{RES}\}}) \geqslant \pi_{it}(T^{*}_{-ijt+ij't}, P^{*}_{-ijt+p_{ij't}^{RES}}) \tag{5}$$

where $T^{*}_{-ijt+ij't}$ refers to the network purchase with good j replaced by good j' for power plant i at time t.

In next section, the abstract inequalities discussed above would be developed into a specific design that is reasonable for estimation in this study.

5 Estimation

In this section, I start with the demand side first, which is standard and thus omitted in the theoretical section. I divide the power plants' demand for coal into two stages. In Stage II, plants determine how much to purchase from each coal inside their purchase set. The purchase set and the total energy requirement are hold fixed in this stage. With a fixed energy requirement, optimizing over quantities is the same as over shares. In Stage I, given the optimal shares, I calculate the overall energy to purchase for fulfilling the electricity generation requirement.

5.1 Estimate the Demand Side

5.1.1 State II – The Nested CES

In each period, power plants purchase from various coal sources to satisfy the energy requirement. A power plant makes continuous choices about the quantities to acquire from various alternatives, which is suitable for a CES-type model. While the standard CES is criticized for the unrealistic IIA property, as standard Logit, Nested CES is a reasonable compromise between flexibility and computational feasibility[33]. CES models the static goals of business operation as constrained optimization. The average revenue from electricity generation is an exogenous process uniformly determined inside a state and beyond the influence of any single power plant. I assume that the goal of plant i is minimizing fuel procurement cost, subject to a budget constraint for the minimum

33. The flexible random-coefficient CES proposed in Adao et al. (2017) would be computationally infeasible here, because I have a computationally very hard supply side.

amount of required energy.

The objective function plant i faces at time t is,

$$\min_{Q_{ijt},\, j\in J_{it}} C_{it} = \left(\sum_{g\in G_{it}} \left[\left(\sum_{j=1}^{J_g} \phi_{ijt}^{1-\rho} Q_{ijt}^{\rho} \right)^{\frac{1}{\rho}} \right]^{\sigma} \right)^{\frac{1}{\sigma}} \quad \text{subject to} \sum_{j=1}^{J_{it}} h_{ijt} Q_{ijt} \geqslant E_{it} \tag{6}$$

C_{it} is the total cost. G_{it} is a set of nests and J_g are total number of goods inside the nest g. J_{it} denotes the purchase set – the set of goods plant i purchase from. Q_{ijt} is the purchased quantity for good j. h_{ijt} is coal's energy content, measured by MMBtu. E_{it} is the energy requirement plant i obligated to meet. $\phi_{ijt}^{1-\rho}$ is the baseline cost level, defined by an exponential function of observed and unobserved characteristics. $\phi_{ijt} = e^{\beta_w w_{ijt} + \beta_p p_{ijt} + \xi_{ijt}}$, with exogenous characteristics w_{ijt}, endogenous variables p_{ijt}. $x_{ijt} = (w'_{ijt}, p'_{ijt})'$. In this case, p_{ijt} is the summation of FOB price and the transportation cost[34]. ξ_{ijt} is the product-individual-time specific error term, unobserved to econometricians, but observable to the power plants and coal companies when prices are negotiated, leading to $Cov(p_{ijt}, \xi_{i\ell t}) \neq 0$. This causes the well-known endogeneity in a structural model. ρ and σ are non-negative nesting parameters.

Given the current energy requirement E_{it}, I transform the optimization over Q_{ijt} into over the energy shares $S_{ijt} = (Q_{ijt} \times h_{ijt})/E_{it}$. Appendix A.1 justifies this simplification. The unconditional S_{ijt} has an analytical solution and can be expressed as a product of $S_{ijt|igt}$ — the choice share for j conditional on g and $S_{igt} = \sum_{j\in J_g} (Q_{ijt} \times h_{ijt})/E_{it}$.

$$\begin{aligned} &\min_{S_{ijt},\, j\in J_{it}} AC_{it} = \left(\sum_{g\in G_{it}} \left[\left(\sum_{j=1}^{J_g} \phi_{ijt}^{1-\rho} \left(\frac{S_{ijt}}{h_{ijt}} \right)^{\rho} \right)^{\frac{1}{\rho}} \right]^{\sigma} \right)^{\frac{1}{\sigma}} \quad \text{subject to} \sum_{j=1}^{J_{it}} S_{ijt} = 1 \\ &S_{ijt} = S_{ijt|igt} \times S_{igt} = \frac{\phi_{ijt} h_{ijt}^{\frac{\rho}{\rho-1}}}{\left(\sum_{m\in J_g} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\left[1-\frac{\sigma(\rho-1)}{\rho(\sigma-1)}\right]} \times \sum_{g\in G_{it}} \left(\sum_{m\in J_g} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\frac{\sigma(\rho-1)}{\rho(\sigma-1)}}} \end{aligned} \tag{7}$$

This transformation from Q_{ijt} to S_{ijt} is crucial to justify the division of Stage I and Stage II, where optimizing over Q_{ijt} is the same as optimizing over S_{ijt} and then have $Q_{ijt} = (S_{ijt} \times E_{it})/h_{ijt}$. Thus, the optimization problem from Stage I and II is solved separately. One can obtain S_{ijt} in

34. The overall price is the price a plant pays to get the coal delivered to its site. Train and Wilson (2011) offers an alternative by treating FOB and transportation costs separately. They came to an unexplained non-rational behavior regarding how plants react to different costs.

Stage II and insert the per-unit cost AC_{it} back to Stage I when calculating E_{it}[35].

To account for this price endogeneity, I follow the equivalence of Nested CES and Nested Logit in Sheu (2014) and the logic behind Berry (1994). After some calculations (Appendix A.1), the system can be inverted as a linear equation of the difference of log S_{ijt} to apply 2SLS.

$$\ln S_{ijt} - \ln S_{ij't} = \begin{cases} (w_{ijt} - w_{ij't})\beta_w + (p_{ijt} - p_{ij't})\beta_p + \xi_{ijt} - \xi_{ij't} + \dfrac{\rho}{\rho - 1}\left[\ln(h_{ijt}) - \ln(h_{ij't})\right], \\ \qquad \text{if } j, j' \in J_{igt} \\ (w_{ijt} - w_{ij't})\beta_w + (p_{ijt} - p_{ij't})\beta_p + \xi_{ijt} - \xi_{ij't} + \dfrac{\rho}{\rho - 1}\left[\ln(h_{ijt}) - \ln(h_{ij't})\right] \\ \qquad + \left(1 - \dfrac{\sigma - 1}{\sigma}\dfrac{\rho}{\rho - 1}\right)\left[\ln(S_{igt}) - \ln(S_{ig't})\right], \quad \text{if } j \in J_{igt}, j' \notin J_{igt} \end{cases} \tag{8}$$

Coals are produced all around the US. The geographic organization remains similar in a specific area, and coals produced there share similar features. Power plants may substitutes towards a similar good in response to a price increase. I divide coal sources into different nests by four major supply regions in the US – PRB, Appalachia, Other Western, and Mid-Continent[36].

Instruments. Gandhi and Houde (2019) recommends constructing instruments using the differences in the products' exogenous characteristics. For good j, I construct the excluded instruments z_{ijt} by calculating the squared and interacted characteristics difference that, $(x_{k,ijt} - x_{k,ij't})^2, \forall j' \in J_{it}$, $\forall k$ and $(x_{k_1,ijt} - x_{k_1,ij't}) \times (x_{k_2,ijt} - x_{k_2,ij't}),\ \forall j' \in J_{it}\ \forall k_1, k_2,$ with $k_1 \neq k_2$.

GMM Objective function. In typical literature of industrial organization, the share of an outside option S_0 usually serves as a non-zero fixed denominator of the inversion. However, heterogeneous power plants purchase from very distinct sources, which prevents the existence of a common

35. This works for the nested CES model. If multiple-discrete continuous choice model is adopted (Train and Wilson, 2011), E_i and the individual purchase Q_{ijt} would not be separable, since in their setting, a higher E_i inducers a larger purchase set.

36. In the Platts data, the coal supply regions are divided into more than 10 small subgroups and I gather them into four major regions. PRB contains what are denoted as Western:Southern PRB, and Western:Northern PRB. Appalachia contains Southern Appalachia, Northern Appalachia:Northeast, Central Appalachia, Northern Appalachia:Ohio. Other Western is composed by other (bituminous) coals produced in the western part, includes Western:Four Corners, Western:Southern Wyoming, Western:Northern Lignite, Western:Washington, Western:Raton/Canon City. Mid-Continent includes Mid-Continent:Illinois Basin, Mid-Continent:Interior, and Mid-Continent:Gulf Lignite.

baseline[37]. Henceforth, I embrace a pairwise difference procedure as in Dubé et al. (2021)[38], and construct the objective function as,

$$\frac{1}{NTJ}\sum_{i=1}^{n}\sum_{t=1}^{T}\sum_{j,j',j\neq j'}\left[\ln\left(\frac{s_{ijt}}{s_{ij't}}\right)-\widetilde{x}_{ijj't}\beta\right]'\widetilde{z}_{ijj't}W\widetilde{z}'_{ijj't}\left[\ln\left(\frac{s_{ijt}}{s_{ij't}}\right)-\widetilde{x}_{ijj't}\beta\right] \tag{9}$$

where $\widetilde{x}_{ijj't}=x_{ijt}-x_{ij't}$, $\widetilde{z}_{ijj't}=z_{ijt}-z_{ij't}$. W is the weighting matrix. s_{ijt} is the share obtained from the data. Note that (9) only use the purchased goods with $s_{ijt}>0$, which implies an assumption that there is no selection on $\xi_{ijt}-\xi_{ij't}$ across the goods with $s_{ijt}>0$ and $s_{ijt}=0$.

One common shortcoming of not having an outside option is that whenever prices of inside goods increase proportionally, consumers would still purchase the same due to the lack of outside choices to deviate to. This drawback is alleviated by Stage I, since Stage I would determine E_{it}. Although $S_{ijt}, j\in J_{it}$ stays the same, a decrease in E_{it} translates into a lower Q_{ijt}.

Other Issues. One factor that needs more attention is plants' on-site stockpiling and the corresponding dynamics. Stockpiling is very common in coal-fired plants. Appendix C.3 discusses this using EIA 902/920 and shows that typical plants store coals amount to 3-month daily usage. Hence, I aggregate the monthly transaction data to quarter-level to reduce its influence.

Another critical issue is the co-existence of long-term contracts, which carry many coal transactions[39]. A long-term contract regularly lasts for two to five years, with possibly one or two modifications in between. Prices agreed in the contract evolve based on the current cost. The dynamics of long-term contracts are hard to capture in this already sophisticated model. Besides, Platts data does not contain enough information to characterize contracts. For instance, I do not observe how the prices would be modified for a contract. Nor do I know a large fraction of the termination date. This drawback impedes and restricts the explicit modeling of contracts. Instead, I have to treat the long-term contracts as a pure exogenous process, affecting and restricting the purchase behavior through current characteristics, prices, and a minimum quantity requirement[40]

37. Even for the world's largest coal mine, Peabody's North Antelope Rochelle, more than half of the coal-fired plants never purchase from it. Besides, not many power plants equip the techniques to switch among different fuel types, precluding using other fossil fuels as an outside choice.

38. If I pick a random good as the denominator for each power plant, the problem is the dependence of GMM objective function value on the good chosen as the baseline. On the contrary, the pairwise difference method pools over all possible combinations to eliminate the dependence on a particular good.

39. Although long-term contracts occupy a vast amount of sales, the price of a newly-signed long-term contract would follow prices in the spot market. Therefore, focusing on the spot market would still be informative.

40. Clauses inside a contract may stipulate the punishment for violations. For instance, if a power plant cannot get the desired quantity supported by a current contract, the coal producer needs to compensate that plant for the price

prevailing in most contracts (Jha, 2019).

5.1.2 Stage I – The reduced form regression

In Stage I, given the average fuel cost AC_{it} in (7) after optimizing over S_{ijt}, a power plant then obtains the E_{it} appears in (6). Briefly, E_{it} is not a choice variable of a power plant. Instead, plants choose S_{ijt} to minimize the AC_{it} and the local electricity distributor chooses E_{it} based on AC_{it}. Furthermore, I model the determination of E_{it} in Stage I through a bunch of reduced-form regression. One advantage of having this Stage I is to resolve the unrealistic substitution pattern without an outside option, as mentioned in Section 5.1.1.

Instead of Stage I, an alternative is having an extra layer of nests in Stage II, indicating the outside option of not purchase. Then both the purchase and the electricity generation is incorporated into the nested CES. This strategy may be problematic since how much electricity to generate may not come from a static optimization. Firstly, the optimal generation under the current transmission grid is not a freely-adjusted variable, and Davis and Hausman (2016) indicates that structurally modeling the electricity grid is prohibitively complicated.

Secondly, the aggregate electricity demand is very inelastic since electricity is necessary for modern society. Without efficient storage techniques, quantities demanded and supplied need to be matched for every short period. Shortage of electricity would cause tremendous property loss for industries and residents. However, shortages due to mismatch would be inevitable if all the power plants made independent one-time decisions about how much to generate in each period. Therefore, regardless of regulation status, power plants need to obey the orders from a centralized planner – possibly the electricity distributor in the area.

From the reasoning above, the total generation for each generating unit inside a power plant is preferably treated as a forecasting problem. Relying on the publicly available high-frequency electricity generation data in CEMS, I estimate a separate reduced-form random effect regression for each generating unit u inside power plant i by,

$$G_{udo} = b_0 + \sum_{j=1}^{B} \left[\mathbf{1}(TG_{do} \in B_j) \times \frac{AC_{it}}{p_{NG,d}} \times b_j \right] + \boldsymbol{CT}_{udo}\boldsymbol{\gamma} + \zeta_{ud} + \epsilon_{udo} \tag{10}$$

gap if the plant has to go to the spot market to fulfill the shortage. This punishment precludes arbitrage behavior when the spot price is high. Conversely, when facing a lower spot price, power plants need to reimburse coal producers for selling in the spot market at a lower price.

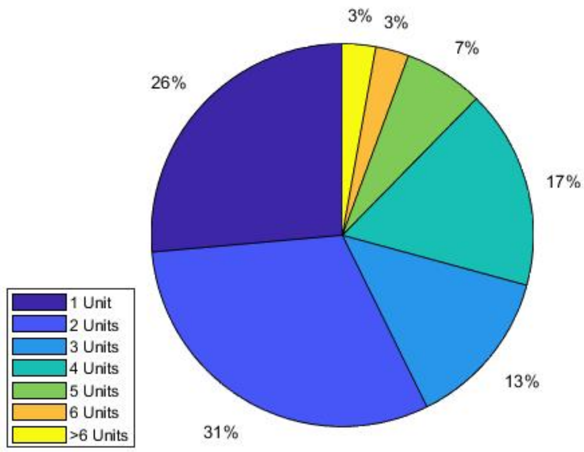


Figure 9: The number of coal-fired generating units a power plant had, during 1998 to 2008. Data Source: CEMS.

G_{udo} is the total electricity generation for generating unit u of plant i at hour o day d. TG_{do} is the total generation in the NERC region plant i belongs to. $AC_{it}/p_{NG,d}$ is the ratio of AC_{it} from Stage II to the real daily natural gas price at the nearest trading hub[41]. b_j are the coefficients over the relative prices of burning coal or natural gas. CT_{udo} are other factors that may influence the total generation[42]. ζ_{ud} is the persistent effect of generating unit u at day d, and ϵ_{udo} is the idiosyncratic error. The total generation within a quarter is the summation of G_{udo}, or $Gen_{it} = \sum_{d \in t} \sum_{o \in d} \sum_{u \in i} G_{udo}$. A power plant may have more than one coal-fired generating unit. Figure 9 displays that approximately 74% of power plants in the data used for structural estimation have more than one coal-fired generating unit.

In equation (10), the relative cost is multiplied by an index function, $\mathbf{1}(TG_{do} \in B_j)$. The total generation at day d hour o in the plant i's NERC region is divided into several bins B_j, interacting with a price coefficient b_j. Thus, I allow for a flexible reaction toward cost fluctuations. Also,

41. I find the nearest trading hub according to the longitude and latitude. Natural gas is transmitted to power plants by a fixed network of pipes. Thus, the location-based distance may not be a perfect proxy for the nearest hub, but it is the best available.

42. Other controls include: hour fixed effect, year-quarter fixed effect, year-quarter fixed effect interact with TG_{do}, weekday fixed effect, some daily characteristics of the total generation – daily maximum, minimum, standard error, and the summation across all the coal-fired generating units in the area. The local temperature data was also incorporated in a previous version because the temperature may influence combustion efficiency. However, in the downloadable temperature record from NOAA https://www.ncdc.noaa.gov/data-access/land-based-station-data/land-based-datasets/global-historical-climatology-network-ghcn, many are missing. Imputation may not be reasonable since daily weather conditions could differ dramatically across stations. Besides, unlike wind power stations, a coal plant is insensitive to the environmental temperature since the combustion process would operate at around 500° to 600°C. Therefore, I do not include temperature as a control.

I include a random effect ζ_{ud} to capture unobserved factors that influence the hourly generation uniformly in a particular day, and ζ_{ud} is assumed to satisfy the conditions for random effect to be consistent. That is, ζ_{ud} is orthogonal to the regressors.

The estimation is performed by the feasible GLS (FGLS). An implicit normalization is necessary to connect the AC_{it} in Stage II and the AC_{it} in Stage I[43]. In Stage I, the estimated AC_{it} from a single individual across different periods is pooled into a reduced-form regression, calling for cross-time normalization for AC_{it}. Without a common outside option, I normalize AC_{it} by finding a set of ξ_{ijt} that the calculated AC_{it} equals to the observed energy-share weighted real price per MMBtu.

In summary, the model in Stage I is a pure forecast reduced-form regression employing high-frequency data. I use the level of generation rather than the log generation because regressing on level results in a higher fit. One advantage of high-frequency data is the unnecessity of accounting for price endogeneity. Plants and coal suppliers negotiate prices every month or quarter so that prices are uncorrelated with the hourly-specific ϵ_{udo}, or the daily-specific ζ_{ud}.

5.2 Estimate the Supply Side

This section illustrates how I turn the network-formation setting in Section 4 into an empirically estimable model. I decompose the abstract surplus function π_{it} into two linear parts, the observed surplus π_{it}^{NFC} and the fixed cost f_{ijt}. As shown in Section 5.1.1, the entire price $p_{ijt} = p_{ijt}^F + p_{ijt}^T$, or the summation of FOB price and transportation cost. Note that taking derivative respect to p_{ijt} is the same as respect to p_{ijt}^F, when transportation cost is treated as exogenous. Therefore, I may use p_{ijt} and p_{ijt}^F interchangeably.

Bargaining Units. Bargaining units describe the entities entering into the Nash bargaining problem. For instance, a large coal company may have several subsidiaries. Most power plants belong to large operating companies[44]. Plants that belong to the same parent company and are close to each other in the distance could be jointly operated. Contrary to this, I ignore the management hierarchy in this study and treat each power plant as an independent decision-maker[45]. On the

43. Similar to the discrete choice model, in the nested CES setting, only the difference, rather than the level, in the error term ξ_{ijt} is identified. A change in the level of $\boldsymbol{\xi}$ would not influence the multiple-continuous result.

44. For instance, Duke Energy, a large company, possesses more than 30 power plants, with regulated utilities for a specific area operating jointly. See https://www.duke-energy.com/Our-Company/About-Us/Power-Plants, and https://www.duke-energy.com/Our-Company/About-Us/Businesses/Regulated-Utilities. The western part of North and South Carolina is operated by Duke Energy Carolina, while Duke Energy Progress operates others.

45. The main reason for this simplification comes from computational tractability. Firstly, allowing for multi-plant operating companies would lead to extra computational burden owing to joint maximization. Secondly, there were

<table>
<tr><th>Company before Merger</th><th>Coal Mines</th><th>Company after Merger</th></tr>
<tr><td rowspan="3">Peabody</td><td>North Antelope Rochelle Complex</td><td rowspan="3">Peabody</td></tr>
<tr><td>Caballo</td></tr>
<tr><td>Rawhide</td></tr>
<tr><td rowspan="4">Rio Tinto</td><td>Antelope</td><td rowspan="4">Rio Tinto</td></tr>
<tr><td>Caballo Rojo</td></tr>
<tr><td>Jacobs Ranch</td></tr>
<tr><td>Spring Creek</td></tr>
<tr><td>Arch</td><td>Black Thunder</td><td rowspan="2">Arch</td></tr>
<tr><td rowspan="2">Triton</td><td>North Rochelle</td></tr>
<tr><td>Buckskin</td><td>Kiewit</td></tr>
<tr><td rowspan="2">Foundation</td><td>Eagle Butte</td><td rowspan="2">Foundation</td></tr>
<tr><td>Belle Ayr</td></tr>
</table>

Table 4: Ownership of the major coal mines in PRB region, before and after the Arch-Triton merger.

supply side, the bargaining unit is a multi-product coal company m. Otherwise, if all the suppliers are considered as single-product, the influence of merger and acquisition could be underestimated. Table 4 summarizes the ownership for major PRB coals before and after the merger. In summary, the bargaining process has the individual power plants bargain with multi-product coal companies in the PRB region over individual goods.

Unlike Ho and Lee (2019), or Ghili (2022), both keeping a small network, this study constructs an extensive network involving all market participants. Across the US, there are around 177 power plants that have the potential to purchase PRB coals. There are twelve major coal mines inside PRB (Table 4). Besides, I observe purchases from other small PRB mines and non-PRB coals. Since the latter is characterized by a much more competitive market with many suppliers, keeping each non-PRB coal as a separate product makes a computationally infeasible problem. To overcome this, I set up two extra composite goods, one for all other PRB coals and the other for all non-PRB coals. Accordingly, I still have $177 \times 14 = 2478$ possible links to check. A tractable analysis for such an extensive network is missing in the current research.

Surplus of Power Plants. The surplus $\pi_{it}(T, P)$ of power plant i at time t is parameterized as

mergers and acquisitions among plant operating companies during the sample periods, bringing about concerns if the operating companies serving as bargaining units.

two parts – a one-time fixed cost of burning coal j at period t[46] and the variable surplus π_{it}^{NFC},

$$\begin{aligned}\pi_{it}(T,P) &= \pi_{it}^{NFC}(T,P) - \sum_{j\in J_{it}} f_{ijt} \\ \pi_{it}^{NFC}(T,P) &= (AR_{it} - \varrho_{it} \times eff_{it} \times AC_{it}) \times Gen_{it}\end{aligned} \tag{11}$$

AR_{it} denotes the average electricity sales revenues for plant i at time t. AC_{it} is the same as the AC_{it} on the demand side. eff_{it} indicates the combustion efficiency, specified as the ratio of energy input (MMBtu) to the electricity generation (MW). ϱ_{it} is the proportion of fuel cost occupied in the general operating cost.

Fixed costs are assumed to be paid by power plants but undertaken by both entities through equilibrium prices. To justify the fixed costs, the generating units inside the old-fashioned coal-fired units may not have enough flexibility. As a result, technical adjustment of operating units is needed to accommodate different coals. Besides, signing a spot contract requires negotiations. Fixed costs also capture the unobservable tastes for plant i to incorporate good j into J_{it}. As mentioned in Section 5.1.1, once J_{it} is finalized, no further selection existis in the nested CES estimation (9).

Moreover, fixed cost is crucial to understanding plants' optimal purchase strategies. In general, they would not purchase all coals available inside the market, which contradicts the prediction of the Nested CES[47]. In this study, this implication is shut down by the fixed cost. The fixed cost may offset the gains from adding an extra good. Therefore, a combination of Nested CES and fixed costs empowers an endogenously-generated realistically not-all-inclusive purchase pattern by balancing the gains of variety with the loss of costs.

FOB price and Transportation cost. In real life, the vertical supply side is more complicated than the duality coal-plant relationship. Except for the mine-mouth plants excluded from analysis, coal mines are usually far from the sites of plants, requesting transportation using railroad, truck, barge, or a combination of these. Transportation is handled by operators independent of coal companies, which make their own pricing decisions. Here I abstract away from the intermediate

46. I do not include any dynamics since I assume plant i encounters this one-time fixed cost f_{ijt} in each period. That is, suppose plant i have already purchased j in t and paid the fixed cost f_{ijt}, in period t + 1, i still needs to pay $f_{ij,t+1}$ to include j inside.

47. A CES-type model would always favor more varieties. The AC_{it} generated by a larger J_{it} of commodities would be smaller than or equal to a restricted set.

transportation operators and treat it as an exogenous process[48]. The transportation costs are known but beyond the control of coal companies during FOB price negotiation, and the FOB prices chosen by power plants are not strategic toward the best response of the transportation operators.

Like the FOB price, transportation cost is only observed for the purchased goods. Since the transportation cost is determined outside the network, I impute them through a reduced-form estimation, $\ln(Trans_{ijt}) = \psi_i + \psi_j + \psi_{1t} + 1(j \in PRB) \times \psi_{2t} + \upsilon_{ijt}$, where the ln transportation cost per mile is assumed to be a linear function of a bunch of fixed effects varied with coal. Because the dependent variable is already per mile, I do not include distance as a regressor. I impute the transportation cost of the observed unpurchased goods using the estimated value of $\widehat{Trans}_{ijt}$.

Then I rewrite the conditions for a stable network in terms of the f_{ijt} and π_{it}^{NFC}.

Condition I. No addition. For a not-purchased good j', plant i would not want to add it into the purchased set since, in equilibrium, adding j' would decrease its overall profit. Mathematically, this would be characterized by an inequality in that the fixed cost surpasses the gains in variable surplus once j' is purchased.

$$\underbrace{f_{ij't}}_{\text{Fixed cost of including j'}} \geqslant \overbrace{\pi_{it}^{NFC}(T^*_{+ij'}, P^*_{+ij'}) - \pi_{it}^{NFC}(T^*, P^*)}^{\text{The surplus gain for i}}, \quad \text{for } Q_{ij't} = 0 \tag{12}$$

The minimum price acceptable to the coal company is the reservation price $p_{ij't}^{RES}$. With coal company m's overall profit denoted by π_{mt}, $p_{ij't}^{RES}$ needs to satisfy,

$$\underbrace{\pi_{mt}(T^*_{+ij'-i'j'}, P^*_{+ij'-i'j'})}_{\text{Profit of selling to i rather than to } i'} \geqslant \overbrace{\pi_{mt}(T^*, P^*)}^{\text{current profits}}, \quad \forall\, Q_{ij't} = 0 \tag{13}$$

With a binding capacity constraint, to sell j' to i, m must switch some quantities of j' currently sold to i'. Otherwise, $i' = \varnothing$. For a single-product company with infinite capacity, $p_{ij't}^{RES}$ would be the marginal cost.

48. A two-sided bargaining model would be more satisfactory in capturing the entire market structure, which leaves for further research. Treating the transportation cost as purely exogenous and unexpected to coal companies could be problematic if a merger induces a different surplus division between coal companies and transportation operators. For instance, the anti-competitive effect of the Arch-Triton merger may be underestimated if, after the merger, the coal companies gain more market power. Then they increase the FOB price to force down the transportation costs.

Condition II. No Cut-off. No cut-off means neither the coal company nor the plant would like to cut an existing solid link unilaterally, since,

$$\underbrace{f_{ijt}}_{\text{Fixed cost of including j}} \leqslant \overbrace{\pi_{it}^{NFC}(T^*, P^*) - \pi_{it}^{NFC}(T^*_{-ij}, P^*_{-ij})}^{\text{The surplus gain for i}}, \quad \forall\, Q_{ijt} > 0 \tag{14}$$

The inequality indicates that the gain from varieties outpaces the loss of the fixed cost of the inclusion of j. Hence, unilaterally cut off the good j would not be a favorable choice for plant i.

Additionally, prices need to be incentive compatible for coal company m, indicated by the inequality in (5.2) below,

$$\overbrace{\pi_{mt}(T^*, P^*)}^{\text{current profit}} \geqslant \underbrace{\pi_{mt}(T^*_{-ij}, P^*_{-ij})}_{\text{Profit after dropping the sales to i}}$$

Condition III. No Replacement. No-replacement condition is equivalent to conditions in which j's price is determined either by the Nash bargaining solution or by the maximum price that m would charge without inducing plant i to replace the j with a current one non-purchased j'. In other words, Nash bargaining solution caps the price for this transaction. Coal companies cannot force plants to accept any price higher than the Nash bargaining price, even when such a higher price is still profitable to i. However, this upper bound may not be attainable with a threat of replacement.

The static Nash bargaining is built on Grennan (2013). Equation (15) describes the problem. p^F_{ijt} is chosen to maximize the product of surplus from both sides, weighted by the bargaining leverage μ, given the current p^T_{ijt}, leading to the overall price $p_{ijt} = p^F_{ijt} + p^T_{ijt}$. J_{imt} contains the set of goods produced by company m and sold to i. $\sum_{\ell \in J_{imt}} Q_{i\ell t}(T^*, P^*)(p^F_{i\ell t} - mc_{i\ell t})$ is the profit obtained by coal company m by selling all the goods in J_{imt} to the plant i, under the current network (T^*, P^*). The disagreement payoff $\sum_{\ell \in J_{imt} \setminus j} Q_{i\ell t}(T^*_{-ij}, P^*_{-ij})(p^F_{i\ell t} - mc_{i\ell t})$ is defined accordingly.

$$\begin{aligned}\max_{p^F_{ijt, j \in J_{imt}}} & \left[\sum_{\ell \in J_{imt}} Q_{i\ell t}(T^*, P^*)(p^F_{i\ell t} - mc_{i\ell t}) - \sum_{\ell \in J_{imt} \setminus j} Q_{i\ell t}(T^*_{-ij}, P^*_{-ij})(p^F_{i\ell t} - mc_{i\ell t}) \right]^{\mu} \\ & \times \left[\underbrace{\left(\pi_{it}^{NFC}(T^*, P^*) - \sum_{\ell \in J_{it}} f_{i\ell t} \right)}_{\pi_{it}(T^*, P^*)} - \underbrace{\left(\pi_{it}^{NFC}(T^*_{-ij}, P^*_{-ij}) - \sum_{\ell \in J_{it} \setminus j} f_{i\ell t} \right)}_{\pi_{it}(T^*_{-ij}, P^*_{-ij})} \right]^{1-\mu}\end{aligned} \tag{15}$$

As illustrated in the intuitive example in Section 2.1 and similar to Grennan (2013), the disagreement payoff is generated by holding all others constant. In other words, if the the bargaining fails for p^F_{ijt}, plant i can purchase as much as it wants to from $J_{it}\backslash j$ at the current price regardless of the capacity constraint in $J_{it}\backslash j$. This is consistent with the single deviation of Nash equilibrium and is also a compromise due to the computational issue – if the network formation is fully absorbed in $\pi_{it}(T^*_{-ij}, P^*_{-ij})$, $\partial\pi_{it}(T^*_{-ij}, P^*_{-ij})/\partial p^F_{ijt} \neq 0$ and the chain reaction would make $\pi_{it}(T^*_{-ij}, P^*_{-ij})$ prohibitively hard to compute.

The difference of (15) from Grennan (2013) lies in the second part. Currently, I put the error term in the fixed costs[49]. Thus, if two goods, both prices derived from Nash bargaining, face the same observables but have different realized prices, this is captured by their differences in fixed costs.

Taking the First Order Condition yields,

$$-\mu \times \frac{\partial\pi_{mt}(T^*, P^*)/\partial p^F_{ijt}}{[\pi_{mt}(T^*, P^*) - \pi_{mt}(T^*_{-ij}, P^*_{-ij})]} = (1-\mu) \times \frac{\partial\pi^{NFC}_{it}(T^*, P^*)/\partial p^F_{ijt}}{[\pi^{NFC}_{it}(T^*, P^*) - \pi^{NFC}_{it}(T^*_{-ij}, P^*_{-ij}) - f_{ijt}]} \tag{16}$$

The bargaining leverage μ divides the surplus between buyers and sellers. For example, $\mu = 1$ eliminates the second part in (16), equivalent to m choosing p^F_{ijt} to maximize its own profits. $\mu = 0$ maximizes the plant's surplus at a minimum price acceptable to company m. In an ideal case, flexible bargaining parameters should be allowed. However, in this study, μ is normalized as $\mu = 0.5$, due to a lack of identification and computational feasibility.

In the supply-side estimation, prices for the purchased goods are observed, and the implied fixed costs are the parameters. Rearrange the items in (16), I obtain an expression of f^{Nash}_{ijt}, which is the implied fixed cost if p_{ijt} is determined entirely by the Nash bargaining,

$$\begin{aligned} f^{Nash}_{ijt} =& \pi^{NFC}_{it}(T^*, P^*) - \pi^{NFC}_{it}(T^*_{-ij}, P^*_{-ij}) \\ &+ \frac{1-\mu}{\mu} \times \frac{[\pi_{mt}(T^*, P^*) - \pi_{mt}(T^*_{-ij}, P^*_{-ij})] \times [\partial\pi^{NFC}_{it}(T^*, P^*)/\partial p_{ijt}]}{\partial\pi_{mt}(T^*, P^*)/\partial p_{ijt}} \end{aligned} \tag{17}$$

The Nash bargaining price in (15) faces the endogenous threat of replacement from no-purchased good j'. j' cannot be a product of company m, because m can eliminate the threat by offering a

49. Literature in empirical Nash bargaining has different frameworks to incorporate the error. For example, Grennan (2013) ascribes the error term in the Bargaining leverages. However, the marginal cost parameters enter nonlinearly into the supply side, causing computational difficulty. This constitutes the parsimonious marginal cost in Grennan (2013). On the contrary, Gowrisankaran et al. (2015) allows the error in the marginal cost and transforms (15) into a linear system regarding to marginal cost instead.

higher price deliberately and instantaneously. A multi-product company would have a competitive advantage since they face fewer threats. A necessary condition for a stable network that is robust to replacement involves one more set of inequalities such that,

$$\overbrace{f_{ij't} - f_{ijt}}^{\text{Loss due to extra fixed cost by replacing j with } j'} \geqslant f_{ijj't}^{diff} = \underbrace{\pi_{it}^{NFC}(T^*_{-ij+ij'}, P^*_{-ij+ij'}) - \pi_{it}^{NFC}(T^*, P^*)}_{\text{Gain in profit when replacing j with } j'} \quad (18)$$

(18) needs to hold for $\forall j, j'$ with $Q_{ijt} > 0$ and $Q_{ij't} = 0$. (18) refers to the case that NFC surplus gain would not fully compensate for the extra loss caused by the differences in fixed costs. No replacement conditions provide inequalities between fixed costs of the purchased and no-purchased goods. For a stable network, the difference in π_{it}^{NFC} should be at least the $f_{ijj't}^{diff}$, under the condition that $p_{ij't}$ is bounded below using the reservation prices $p_{ij't}^{RES}$ implied following,

$$\underbrace{\pi_{mt}(T^*_{+ij'-i'j'}, P^*_{+ij'-i'j'})}_{\text{Profit after switching to selling to i from } i'} \geqslant \overbrace{\pi_{mt}(T^*, P^*)}^{\text{current profits}} \quad (19)$$

Capacity constraints. In this study, I assume coal companies face capacity constraints and cannot supply as many quantities as they want. The productive restrictions do not hail capacity constraints since coal mines generally operate below their maximum mining capability[50]. Instead, two other factors confine the trading volumes. The first is the capacity-constrained railroad connecting PRB with the rest of the nation. Kaplan (2007)'s report on railroad reliability mentions that, since the mid-1990s, the tight capacity restriction has become part of the reasons for delays in coal deliveries. The second restriction is the mine-waste-disposal limit regulated by EPA. For example, coal mines must apply for permits to discharge the waste into local water. Mines would quit producing when the aggregated waste disposal exceeds the quantities stipulated in the applied permit.

Capacity constraints enter into the model through p_{ijt}^{RES} and prevent p_{ijt}^{RES} from dropping to its marginal cost. When a capacity constraint is restrictive, the cost of selling good j to i becomes the opportunity cost, characterized by the maximum markup company m obtains by selling to i' instead. p_{ijt}^{RES} is the minimum acceptable price that would not induce j's single deviation towards selling to i' instead. Usually, this opportunity cost is higher than the marginal cost. If an existing capacity constraint is ignored, one may have fixed costs estimated abnormally high for non-purchased goods

50. https://www.eia.gov/todayinenergy/detail.php?id=46096 illustrates the productive constraint. There were always spared capacities.

since only a high fixed cost can offset the surplus from purchasing at marginal cost in an oligopolistic market[51].

Calculate the p_{ijt}^{RES}. Capacity constraints enter the model through the reservation price p_{ijt}^{RES}. In the illustrative example in Section 2.1, all the dashed lines are associated with observed prices, and those prices, like $P_A(2)$ or $P_B(1)$, serve as the price a supplier would offer a buyer. However, this is not compatible with real data.

As mentioned in Section 2.2, one problem with the individual transaction data is the partially-observed prices. The proposed prices are not recorded unless the transaction happens, which must be imputed if missing. Here I impute the unobserved prices with p_{ijt}^{RES}, the minimum price acceptable for company m to sell to i rather than i'. p_{ijt}^{RES} depends on the current network and capacity constraints. As discussed in previous sections, without capacity restrictions, $i' = \varnothing$, and the minimum price acceptable to a single-product supplier m is the marginal cost. Binding capacity constraints complicate the calculation. Appendix B.1 fully illustrates algorithms. Briefly, under a given network (T, P), Appendix B.1 solves p_{ijt}^{RES} as a block-wise optimization using a greedy algorithm.

Parameterize the f_{ijt}. Fully nonparametric point identification of f_{ijt} is impossible. Instead, I parameterize the f_{ijt} as a linear function of some observed characteristics Ω_{ijt}, product fixed effects FE_j^r, and time fixed effects FE_t^r.

$$f_{ijt} = FE_j^r + FE_t^r + \chi\Omega_{ijt} + \nu_{ijt} \tag{20}$$

The superscript r $\in R$ in equation (20) denotes the plant's NERC region, and the coefficients with a superscript r vary with those regions[52]. ν_{ijt} is the error term, capturing the variation in f_{ijt} that

51. Reconciling a binding capacity constraint with the Nash bargaining solution in (15) desires further thinking. Typical works like Grennan (2013) re-optimize over the rest of goods in the purchase set through an unconstrained manner, holding all else constant. If the capacity constraint is carved in stone for each purchased good separately, this re-optimization for disagreement payoff is not allowed. Nevertheless, suppose the capacity constraint is either slack or, if it is more about a transportation restriction on the total quantities of coal conveyed out of PRB, then the re-optimization is justified. The quote of Q_{ijt} would be assigned to other goods when the bargaining over j fails. Since the nested CES structure leads power plants to favor diversity, the loss of one good would increase the implied average variable cost and result in fewer total quantities consumed by the plant. Usually, I have $Q_{ijt} \geqslant \sum_{j' \in J_{it}} Q_{ij't}(T^*_{-ij}, P^*_{-ij})$.

52. Because the NERC division changes over time, I aggregate some of them towards five big regions – (i) SPP + ERCOT-TRE, (ii) MRO, (iii) SERC + MAIN-SERC, (iv) WECC, (v) ECAR-RFC. Other fixed effects, like the power plant level fixed effect, FE_i^r, would make it a hard convergence numerically. Thus I focus on a relatively parsimonious regression without FE_i^r.

departs from the linear equation.

Parameterize the mc_{ijt}. Searching for the marginal cost function is computationally challenging. Analysis below illustrates that the marginal costs for PRB coals are nonlinear parameters, which compose the major computational difficulty for this model. Accordingly, similar to Grennan (2013), I assume a parsimonious marginal cost function for the PRB coal with

$$mc_{ijt} = \alpha_{PRB} \times \mathbf{1}\{j \in PRB\} \tag{21}$$

where the marginal cost equals an average value α_{PRB} for all the PRB coals[53]. Besides, the marginal cost of the non-PRB coals is not identified because the non-PRB coal enters as an exogenous composite good, with prices staying at their observed l

Minimum purchase obligation for long-term contracts. By exploring the historical data, Jha (2019) points out that a minimum periodic purchase is prevailing among contracts. A complete model needs to account for this; otherwise, the purchase behavior in the spot market may be too flexible in response to a structural change. I include the minimum purchase requirement through a penalty in π_{it}^{NFC},

$$\pi_{it}^{NFC}(T, P) = \overbrace{(AR_{it} - \varrho_{it} \times eff_{it} \times AC_{it}) \times Gen_{it}}^{\pi_{it}^{NFC}(T,P) \text{ in } (11)} - \underbrace{\sum_{J_{it}} \{\max\{0, \underline{Q}_{ijt} - Q_{ijt}\} \times p_{ijt}^{F}\}}_{\text{Penalty if violation}} \tag{22}$$

The penalty is composed of the positive difference between the required quantity $\underline{Q}_{ijt}$ and the actual purchase Q_{ijt}. $\underline{Q}_{ijt} = 0$ if j is a spot good, but is set as the 90% of the minimum observed quarterly quantity purchased if j is a long-term contract. Non-compliance of $\underline{Q}_{ijt}$ causes extra coal stocks in j for company m, and I assume that for every positive $\underline{Q}_{ijt} - Q_{ijt}$, plant i has to compensate the coal company with the spot price. Furthermore, equation (22) knocks down the monotonicity of $\pi_{it}^{NFC}(T, P)$ regarding p_{ijt}. For plants holding long-term contracts, there is a lower bound of p_{ijt}, below which the penalty term dominates. Thus, there may not be a Nash bargaining price that satisfies the FOC in (16), which would be problematic only in counterfactual [54]. In counterfactual

53. Certainly, this parsimonious setting is not realistic. I would expect some diminishing, increasing return to scale, or a combination in a production function. Moreover, different coal companies may have distinct cost levels. (21) abstracts away from all these differences.

54. The minimum purchase obligation is defined as 90% of the observed purchase, which would never bind in realized data.

calculation, I handle the case where there is no Nash bargaining price solving (16) as a bargaining failure and remove the link in the network.

Timeline. At time t, with the current capacities held on by long-term contracts, power plants and coal companies enter the spot market. They build a trading relationship with each other by searching for a stable network, conditional on knowing the complete information of plants' optimization in (7) and (10). The decision variable is only price, not quantity. Quantities are determined by the optimization of power plants, given the prices. They hypothetically add, cut down and replace the links in the network until everyone in the market are unilaterally satisfied. For each network, the hypothetical equilibrium prices are calculated. After obtaining a stable network, buyers and sellers come together and trade according to the pre-determined "menu" for a limited period. Those prices and trading networks are acceptable to both sides since the network in the "menu" is stable. Buyers would accept the prices because their deviation would not lead them to better places.

Simultaneously, buyers may also order coal delivery through long-term contracts they currently hold. Once the transaction is accomplished, the relationship built on the spot market dissolves. Both sides wait for the next period to enter the spot market again.

The overall optimization problem. Gathering inequalities (12), (14) and (18), and equations (17), the equilibrium in supply side can be summarized into a high-dimensional constrained nonlinear optimization problem ($*$) such that,

$$\min_{f_{ijt},\nu_{ijt},FE^r_j,FE^r_t,\chi,\alpha_{PRB}} \quad \sum_{i,j,t} \nu^2_{ijt}$$

$$\textbf{s.t.} \quad f_{ijt} = FE^r_j + FE^r_t + \chi\Omega_{ijt} + \nu_{ijt}, \ \forall i,j,t, \ r \in R$$

$$f_{ij't} \geqslant \underline{f}_{ij't}, \quad \forall j' \text{ with } Q_{ij't} = 0 \tag{23}$$

$$f_{ijt} \leqslant \overline{f}_{ijt}, \quad \forall j \text{ with } Q_{ijt} > 0 \tag{24}$$

$$f_{ijt} = \{f^{Nash}_{ijt}, f_{ij't} - f^{diff}_{ijj't}\}, \ \forall Q_{ij't} = 0, \ \forall Q_{ijt} > 0 \tag{25}$$

$$\begin{aligned} f^{Nash}_{ijt} &= \pi^{NFC}_{it}(T^*,P^*) - \pi^{NFC}_{it}(T^*_{-ij},P^*_{-ij}) \\ &\quad + \frac{1-\mu}{\mu} \times \frac{[\pi_{mt}(T^*,P^*) - \pi_{mt}(T^*_{-ij},P^*_{-ij})] \times [\partial\pi^{NFC}_{it}(T^*,P^*)/\partial p_{ijt}]}{\partial\pi_{mt}(T^*,P^*)/\partial p_{ijt}} \end{aligned} \tag{26}$$

$$\underline{f}_{ij't} = \pi^{NFC}_{it}(T^*_{+ij'},P^*_{+p^{RES}_{ij't}}) - \pi^{NFC}_{it}(T^*,P^*) \tag{27}$$

$$\overline{f}_{ij't} = \pi^{NFC}_{it}(T^*,P^*) - \pi^{NFC}_{it}(T^*_{-ij},P^*_{-ij}) \tag{28}$$

$$f^{diff}_{ijj't} = \pi^{NFC}_{it}(T^*_{-ij+ij'},P^*_{-ij+p^{RES}_{ij't}}) - \pi^{NFC}_{it}(T^*,P^*) \tag{29}$$

In parallel with Ghili (2022), the objective function in ($*$) is to minimize the variance of the error term of f_{ijt}, subject to f_{ijt} follows a linear additive equation and satisfies a bunch of inequalities and nonlinear equalities for a stable network. ν_{ijt} is a linear function of f_{ijt} and can be partitioned out immediately. This setting differs from an ordinary regression with f_{ijt} as the dependent variable since f_{ijt} is not observed. The optimization is a high-dimensional mixed-integer nonlinear optimization, with the mixed-integer introduced through the equality constraint (25), and the nonlinear comes from the calculation of $p^{RES}_{ij't}$. (25) refers to the observed prices pinned down either by Nash bargaining or the lowest upper bound implied by the no-replacement.

(25) in ($*$) is described as a choice among a set of discrete options implied in the model. With $\mu = 0.5$, the Nash bargaining price is less than the monopolistic price. Thus, profit-optimization sellers have no incentive to undercut the minimum of the Nash bargaining and no-replacement prices, which departs from the quadratic optimization framework in Ghili (2022).

Unfortunately, with (25), ($*$) fails to be a convex optimization and suffers from computational complexity. Here I grasp a computational gain by partitioning out f_{ijt}, $Q_{ijt} > 0$ and replacing (25) by

Subset	Coefficients	Explanation
Nonlinear Coefficients	α_{PRB}	Coefficients entering into the nonlinear calculation of f_{ijt}^{Nash}, $\underline{f}_{ijt}$, and $f_{ijj't}^{NFC}$, and do not appear anywhere else after those three are calculated.
Linear Coefficients	$f_{ij't}$ with $Q_{ij't}=0$, FE_j^r, FE_t^r, χ.	Can be relatively high-dimensional. NOT entering into the determination of f_{ijt}^{Nash}, $\underline{f}_{ijt}$, and $f_{ijj't}^{NFC}$. Could be calculated once f_{ijt}^{Nash}, $\underline{f}_{ijt}$, and $f_{ijj't}^{NFC}$ are given.
Partitioned-out Coefficients	f_{ijt} with $Q_{ijt}=0$ ν_{ijt}	Can be calculated following deterministic functions for the current iteration of nonlinear and linear coefficients.

Table 5: Coefficients calculated in the supply side.

$f_{ijt} = \min\{f_{ijt}^{Nash}, \min_{j' \text{ with } Q_{ij't}=0}\{f_{ij't} - f_{ijj't}^{diff}\}\}$, $\forall j$ with $Q_{ijt} > 0$ [55]. An underlying equilibrium selection mechanism is behind this. Specifically, if f_{ijt}^{Nash} satisfies all no-replacement conditions, there is $f_{ijt} = f_{ijt}^{Nash}$. This selection mechanism is equivalent to the Nash bargaining price is the maximum price that coal companies can force the power plants to accept if π_{it}^{NFC} regardless of the minimum quantity restrictions from long-term contracts. A rough proof of this can be found in Appendix B.2.

In $(*)$, the coefficients to estimate are $f_{ijt}, \nu_{ijt}, FE_j^r, FE_t^r, \chi, \alpha_{PRB}$. A close inspection shows that, after partition-out, these coefficients left can be divided into two subsets based on the way they enter into $(*)$. Table 5 gives an explanation of different types of parameters. Concisely, nonlinear parameters like α_{PRB} are parameters that determine $(*)$, and linear parameters are those solved by the Inner Loop in Algorithm 1 for a given value of nonlinear parameters.

Computational Algorithm. The optimization problem $(*)$ has a nested structure and can be solved with an inner-loop outer-loop iteration. The outer loop optimize the nonlinear parameter α_{PRB} that enters into the calculation of f_{ijt}^{Nash}, $\underline{f}_{ij't}$, and $f_{ijj't}^{diff}$. During this, one also needs to calculate $p_{ij't}^{RES}$, $\forall j'$ with $Q_{ij't} = 0$. The inner loop searches over linear parameters $f_{ij't}$ for $Q_{ij't} = 0$, FE_j^r, FE_t^r and χ to minimize the variance of the error term ν_{ijt} given f_{ijt}^{Nash}, $\underline{f}_{ij't}$, and $f_{ijj't}^{diff}$, and (20). The feasibility of this algorithm derives from the separability of the calculation of f_{ijt}^{Nash}, $\underline{f}_{ij't}$, and $f_{ijj't}^{diff}$ from f_{ijt}. To put it in another way, when calculating f_{ijt}^{Nash}, $\underline{f}_{ij't}$, and $f_{ijj't}^{diff}$, the exact level of f_{ijt} is not needed.

Though having a similar nested structure as Berry et al. (1995), problem $(*)$ is more computa-

55. A recent research, Dedieu et al. (2021) shows that cyclic coordinate descent is applicable to unconstrained mixed integer problem also. So this may not be necessary.

tionally challenging. Firstly, in this case, the outer loop only gives constraints rather than the exact f_{ijt} for the inner loop. The inner loop does not have an analytical solution still. Besides, f_{ijt} are connected explicitly by the linear coefficients and cannot be solved market by market independently. Consequently, a simultaneous nonlinear search over all fixed costs of the non-purchased goods may be computationally impractical. To settle this, I refer to the Alternating Optimization algorithm, also known as Block Coordinate Descent, where the recent development includes Hong et al. (2016). Algorithms like Hong et al. (2016) iteratively decompose a large infeasible problem into many easily-managed small optimizations.

Algorithm 1 shows the exact computational procedure.

Algorithm 1. *An iterative algorithm to solve for problem* $(*)$ *proceeds on in the following steps:*

Outer Loop: *Search over* α_{PRB}. *For the current* α_{PRB}, *calculate* f_{ijt}^{Nash}, $\underline{f}_{ij't}$, *and* $f_{ijj't}^{diff}$. *With capacity constraints, obtain* $p_{ij't}^{RES}$ *through optimizations in Appendix B.1.*

Inner Loop:

i). *Given the current guess of* $\boldsymbol{f}_{ij't}^{k}$, *solve for the linear parameters* χ^k, FE_j^k, *and* FE_t^k *by the OLS on* $f_{ijt} = FE_j^k + FE_t^k + \chi^k\Omega + \nu_{ijt}$.

ii). *Solve for the new iteration* $\boldsymbol{f}_{it}^{k+1}$, *market by market, by optimizing,*

$$\begin{aligned} &\boldsymbol{f}_{it}^{k+1}, \boldsymbol{\nu}_{it}^{k+1} = argmin_{f_{ijt},\nu_{ijt}} \sum_{j \in Market\ i\text{-}t} \nu_{ijt}^2 \\ &s.t.\ f_{ijt} = FE_j^k + FE_t^k + \chi^k\Omega_{ijt} + \nu_{ijt},\ \forall j \\ &\quad f_{ij't} \geqslant \underline{f}_{ijt}, \quad \forall j'\ with\ Q_{ij't} = 0 \\ &\quad f_{ijt} = \min\{f_{ijt}^{Nash}, \min_{j'\ with Q_{ij't}=0}\{f_{ij't} - f_{ijj't}^{diff}\}\},\ \forall j\ with\ Q_{ijt} > 0; \end{aligned} \tag{30}$$

This optimization shown in (30) is a manageable small problem restricting to an individual market i-t, conditional on the fixed FE_j^k, FE_t^k, *and* χ^k. *Here I just omit the superscript of region r for simplicity.*

iii). *Repeat* ***i)****, an OLS regression with an analytical solution, and* ***ii)****, a bunch of small-scale nonlinear optimization problems, each can be solved speedily until converge or some termination conditions are met.*

iv). *The objective function value corresponding to* α_{PRB} *is* $\sum_{i,j,t} \nu_{ijt}^2$ *evaluated at the optimal solution from* ***iii).*** *Search over* α_{PRB} *until finding a minimum of* $\sum_{i,j,t} \nu_{ijt}^2$.

Further discussion about the details of this algorithm can be found in Appendix B.2. It turns out that Algorithm 1 is downhill descending and would stop at a fixed point. That fixed point is a local minimum[56].

6 Estimation Results

This section displays the estimated results from the structural model, including the demand-side Stage II, Stage I, and the supply side. In estimation, I drop the periods from 2005Q3 to 2006Q2 due to the railroad derailment in the PRB. This long-lasting malfunction in transportation emptied plants' coal storage. It also made the plants care less about prices so long as to keep an uninterrupted coal supply, which may contaminate the demand estimation.

6.1 Demand Side Estimation

Demand side – Stage II. The estimated coefficients for the demand side, Stage II, are presented in Table 6. Long-term contracts and spot purchases are pooled together. Products' characteristics include sulfur content, ash content, sulfur content interacting with the installation of scrubbers, delivery prices, distance, regulation status, and all interacting with the long-term contracts indicator. I have attempted several settings where most of the coefficients stay similar. The price coefficients are negative across all settings, and becomes significant when controls for long-term contracts added in[57]. Plants still care about the log distance (in thousand miles) even when the transportation cost is accounted for in the prices. Deregulation does not alter price sensitivity. Based on the Sargan p-value from the over-identification test. I will stick to Model 4 from now on.

The standard errors for the nesting parameters σ and ρ are obtained through the delta method. In Model 4, both of σ and ρ are larger than 1 and $\hat{\sigma} > \hat{\rho}$, which indicates a reasonable nesting structure. Figure 10 shows the implied price elasticities of demand. Although the fuel demand as a

56. Usually, a non-convex problem may not have a unique minimum. For non-convex optimization, no algorithm is guaranteed to converge into the global minimum unless initializing over many starting points. Hence, Algorithm 1 at least does not perform worse than any typical non-convex optimization.

57. Although (6) is a cost minimization problem, I still expect the price coefficient to be negative, because the baseline cost ϕ_{ijt} enters into (6) with power $1 - \rho$ and $\rho > 1$.

whole for power plants is inelastic[58], there can still be moderate to high elastic when the quantity demanded is restricted to a particular good. Own price elasticities are larger in absolute value for the spot compared to long-term purchases. There is more substitution inside the nest rather than across, though the difference is not large.

Demand side – Stage I. In Stage I, the forecast-type reduced-form regression in (10) is carried out. A separate regression is applied to each generating unit u, leading to a bunch of estimated coefficients, which are hard to display as in Table 6. Alternatively, Figure 11 depicts the distribution of estimated coefficients regardless of significance through a panel of histograms.

Generally, the power plant's hourly generation is insensitive to the relative price change. There are not many dissimilarities among different TG_{do} levels in the NERC region, except for a significant fraction of b_j becomes negative when TG_{do} increases. At first sight, this result may be counter-intuitive since one may expect the generating units operating to fulfill the demand at rush hours (large TG_{do}) to be less sensitive to fuel costs. Nevertheless, a large TG_{do} is usually accompanied by a large G_{udo}, giving room for fluctuations in electricity generation to step in. Put differently – plants may keep a stable baseline generation for foundational combustion efficiency regardless of fuel costs, yet the extra capacity tends to be assigned to plants with some cost advantage.

Besides, I need to emphasize that the b_j is the coefficient in front of $\mathbf{1}(TG_{do} \in B_j) \times (AC_{ijt}/P_{NG,d})$, or the ratio of price per MMBtu[59]. For most generating units, b_j is negative, meaning that an increment of the expenses in burning coals relative to the Natural Gas leads to less coal-fired electricity generation. However, b_j could be positive for a few generating units. Several potential reasons apply to this positivity. Firstly, generating units inside a power plant could be heterogeneous in cost-efficiency, and plants may re-optimize over different units in response to fuel price changes. Secondly, a similar situation may happen when distributors assign generation quotes across plants.

58. See, for instance, the EIA report here https://www.eia.gov/analysis/studies/fuelelasticities/pdf/eia-fuelelasticities.pdf

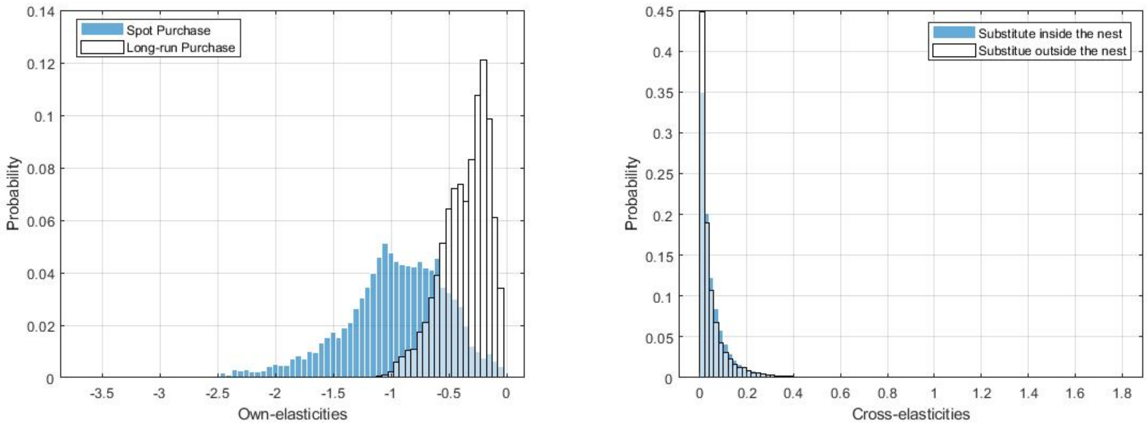
59. At first glance, the considerable absolute value of b_j could be misleading. However, the b_j is the marginal effect of the quotient of the average cost AC_{ij} optimized in Stage II with $P_{NG,d}$. For instance, the total generation would change by b_j only when AC_{ij} doubles for coal but $P_{NG,d}$ unchanged for natural gas, which is almost impossible, especially when AC_{ij} is the cost-minimization solution.

	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4	(5) Model 5
Price	-0.0093	-0.0403***	-0.0466***	-0.0388***	-0.0507***
	(0.0145)	(0.0125)	(0.0166)	(0.0145)	(0.0133)
Sulfur Content	-0.0083	-0.0159***	-0.0178***	-0.0050	-0.0117***
	(0.0060)	(0.0042)	(0.0053)	(0.0053)	(0.0045)
Ash	-0.0379***	-0.0560***	-0.0560***	-0.0406***	-0.0424***
	(0.0034)	(0.0039)	(0.0052)	(0.0051)	(0.0051)
Sulfur × Scrubber	0.0576***	0.0481***	0.0476***	0.0454***	0.0451***
	(0.0042)	(0.0051)	(0.0053)	(0.0052)	(0.0052)
Price × LR		-0.0017	-0.0016	0.0216***	0.0170***
		(0.0014)	(0.0021)	(0.0048)	(0.0045)
Sulfur × LR		-0.0100***	-0.0110**	-0.0195***	-0.0197***
		(0.0034)	(0.0047)	(0.0040)	(0.0040)
Ash × LR		0.0622***	0.0625***	0.0161	0.0269***
		(0.0045)	(0.0065)	(0.0101)	(0.0094)
SO2 × Scrub × LR		0.0094**	0.0100**	0.0207***	0.0206***
		(0.0040)	(0.0046)	(0.0045)	(0.0046)
Price × De-regu			0.0074		
			(0.0386)		
Distance				0.9560**	0.2242
				(0.3913)	(0.2078)
Distance × LR				-0.9639***	-0.7328***
				(0.1939)	(0.1730)
ρ	2.2832*	1.6644***	1.5864***	1.8369**	1.6231***
	(1.1679)	(0.3569)	(0.3453)	(0.7222)	(0.2750)
σ	4.9400***	1.8007***	1.6487***	2.8962***	-
	(1.2334)	(0.3592)	(0.3267)	(0.8558)	-
Date-Region-FE	Yes	Yes	Yes	Yes	Yes
Date-Region-LR-FE	Yes	Yes	Yes	Yes	Yes
N	232362	232362	232362	232362	232362
Sargan p-value	0.9186	0.0002	0.0000	0.7375	0.0085
GMM C Statistics	0.0051	0.0000	0.0000	0.1068	0.0684

Standard errors in parentheses. Clustered at plant-level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 6: Second Stage Demand – Nested CES Demand Estimation



(a) Own price elasticities

(b) Cross price elasticities

Figure 10: The price elasticities for the demand side, stage II.

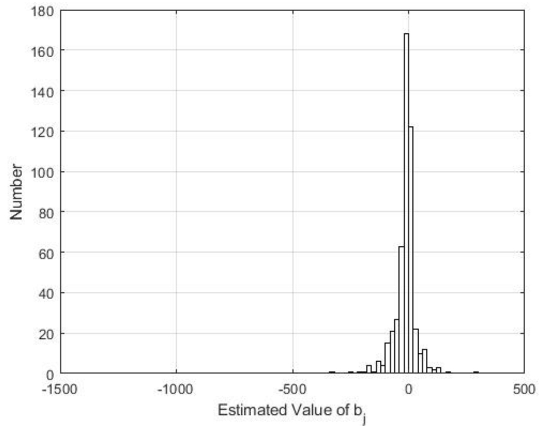
6.2 Supply Side Estimation

The supply side estimation is implemented by combining a fine grid search over nonlinear parameter α_{PRB} and a nonlinear optimization overall other linear parameters. The motivation for this compromise is a pure computational issue. A numerical optimization over a continuous value of α_{PRB} is not computationally feasible because of the magnitude of the network I have[60].

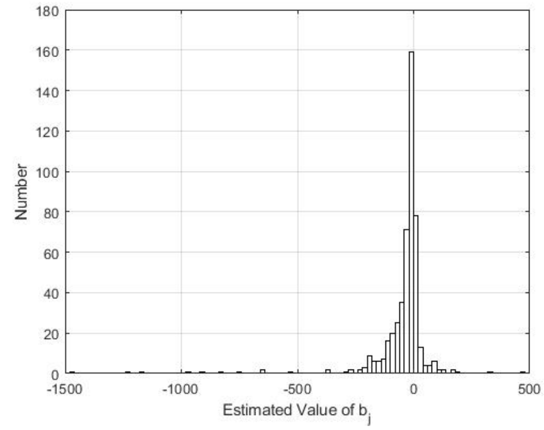
There are few observed characteristics determining the fixed cost f_{ijt}, entering into the model linearly through $\chi\Omega_{ijt}$ in (20). $\Omega_{ijt,1}$ is the proportion of coal-fired generating units with a scrubber installed at time t. $\Omega_{it,2}$ is an indicator equal to 1 if plant i currently holds a long-term contract from the good j's producer. $\Omega_{it,3}$ is the total number of long-term contracts held contemporarily, a proxy of plant i's preference of obtaining fuels from the spot market. This is also an approximate measure of the plant i's unobserved tastes – whether or not plant i is comfortable purchasing from various sources.

The supply side's estimated non-fixed-effect and non-fixed-cost coefficients are displayed in Table 7. I estimate two settings motivated by different capacity constraints. Setting 1 has a tight capacity constraint, where I assume the observed sales quantities are the same as PRB coal's production constraints. On the other hand, setting 2 implies that coal mines never exhaust their production capabilities. While the exact value of capacity constraints is not publicly available, I use the type-

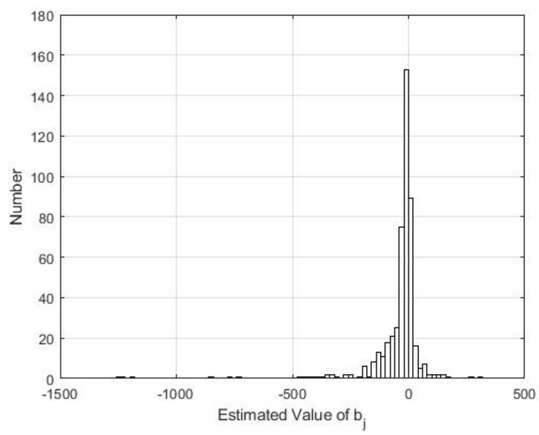
60. The structural model itself does not impede any estimation. For a smaller network in Ho and Lee (2019), even exhausting all possible networks is practicable.



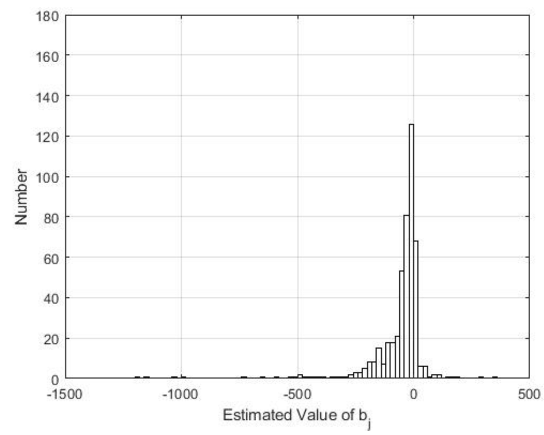
(a) j = 1, the lowest TG_{do}.



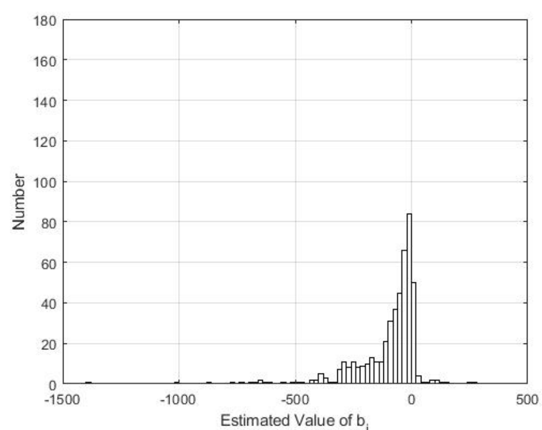
(b) j = 2



(c) j = 3



(d) j = 4



(e) j = 5, the highest TG_{do}.

Figure 11: The distribution of the estimated b_j, separately drawn for different levels of TG_{do}.

aggregated data as an approximate alternative[61]. If the sales quantities are bounded above by other factors rather than the productive ability, like the maximum volume of railroad shipment discussed in Section 5.2, setting 1 would be preferable. In the estimation and the counterfactual, I impose capacity restrictions by soft penalties, similar to the minimum quantity requirement for contracts in (22). I assume a skyrocketing marginal cost whenever there is violation of capacity constraint, which drives the supplier away from the current situation.

I search α_{PRB} over a fine grid inside a closed interval[62]. Figure 12 shows the objective function value $\sum_{i,j,t} \nu_{ijt}^2$ for all the candidate points of α_{PRB} in the fine grid, which takes a irregular shape but at least is has a global minimum solution. Table 7 displays that, as expected, if plant i currently holds a contract from the same coal company, then plant i would encounter a lower fixed cost since they have already become a close partner. Transaction between them is cost-saving. For χ_3, comfortable burning coals from different sources in long-term contracts translates into similar patterns in the spot market. Moreover, χ_1 is positive since plants with scrubbers installed are less connected to PRB coal. Given the same price level, PRB coal is less attractive, which may drive up the f_{ijt} through the inequality constraint (23).

Besides those structural coefficients χ, I calculate a full set of fixed costs f_{ijt}, $\forall j$, whose distribution is displayed in Figure 14. Fixed costs are smaller for the purchased goods than the non-purchased goods by construction[63]. Interestingly, in setting 1, the average of f_{ijt} increases after the Arch-Triton merger, indicating that some plants may find the PRB coal less favorable after the merger, or, more likely, after all the unpredictability caused by the train derailment starting from May 2005. The magnitude of the differences decreases in setting 2, though I still have a slightly larger after-merge average.

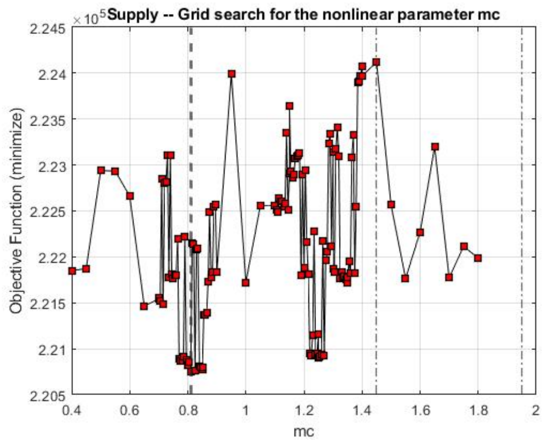
	α_{PRB}	χ_1	χ_2	χ_3
Setting 1	0.8100	0.2596	-2.8700	-0.2746
Setting 2	0.9300	0.1870	-2.6983	-0.2367

Table 7: The obtained coefficients from find grid search.

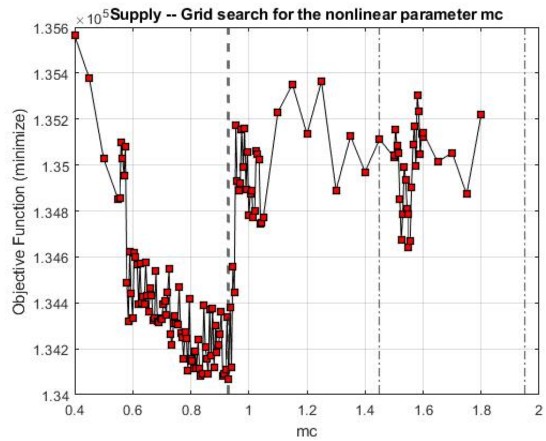
61. The data on the yearly type-specific capacities can be found in https://www.eia.gov/todayinenergy/detail.php?id=46096. All PRB coals belong to the surface-mining category. The capacity usage varied across years, and around 1/6 to 1/5 of the total production capacity was spared.

62. A stable network does not allow selling products at prices below the marginal cost. Then one drawback of this parsimonious cost structure is that α_{PRB} is restricted above by the smallest FOB price observed in the data. Fortunately, I do not have a boundary solution in any settings.

63. I do not restrict the fixed costs for the purchased goods, so they are allowed to drop below 0

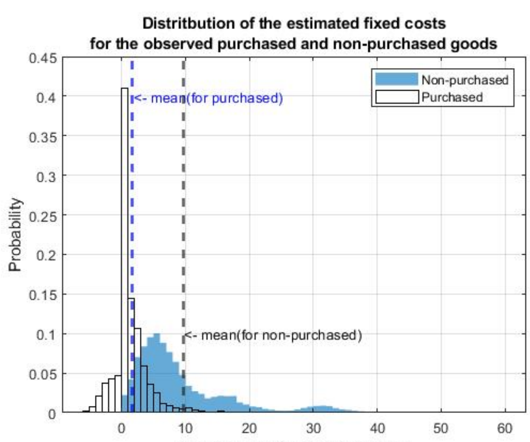


(a) Setting 1

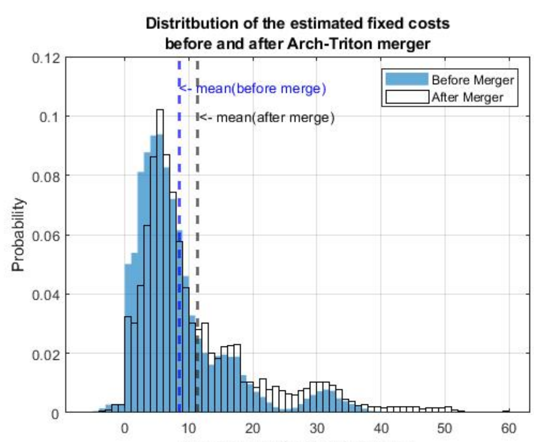


(b) Setting 2

Figure 12: Objective function value corresponding to α_{PRB} regarding two different settings.



(a)



(b)

Figure 13: Distribution of the estimated fixed costs, Setting 1.

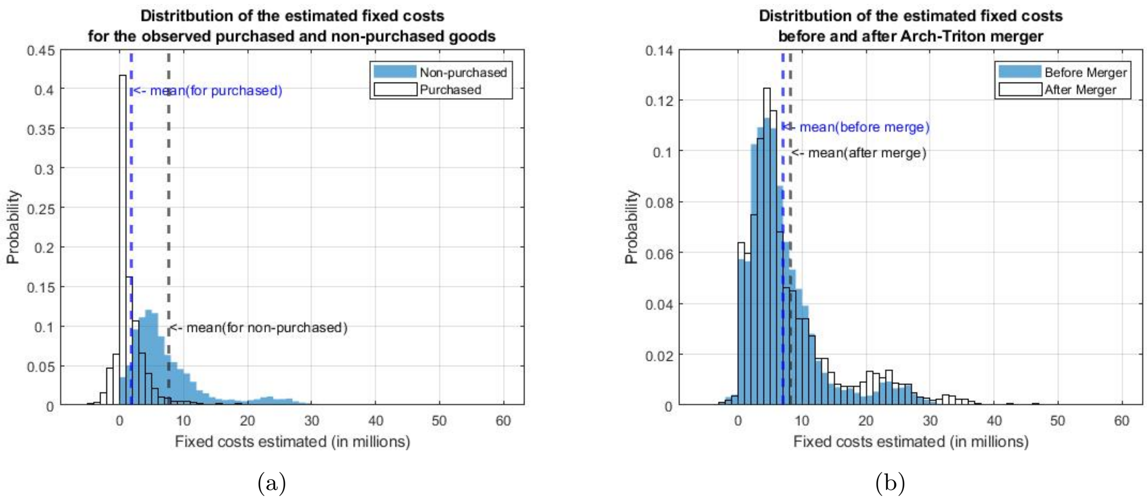


Figure 14: Distribution of the estimated fixed costs, Setting 2.

7 Counterfactual Analysis

The objective of the counterfactual algorithm I have here is to compute the new equilibrium stable network (T^*,P^*) after a hypothetical structural change. Unlike the small number of the possible networks in Ho and Lee (2019), the network between coal suppliers and plants is too large to be traversed. With 14 different goods on the supply side and more than 170 individual plants searching for purchasing in the spot market, one needs to check around $2^{14\times170}$ different networks to find a global solution.

To circumvent this computational infeasibility, I modify the heuristic method proposed in Ghili (2022) and Fan and Yang (2020) to allow for a more flexible design. I restrict the final solution inside the set of networks that can be arrived from the starting point through a series of link additions and cut-offs. From the beginning, a sequence of individual actions is taken in a specific order until no more such unilateral actions are incentive compatible. Then the system attains a new equilibrium or a different stable network.

Fan and Yang (2020) demonstrate that their heuristic method can recover reasonable results in their counterfactual simulations. And they show that these one-step deviations still can generate a distinct pattern despite several drawbacks. For instance, the solution it achieves need not be a global optimum and may be very distinct from the maximum social-welfare results after traversing all networks in Ho and Lee (2019). More likely, it would stop at a network close to where it starts[64]. Besides, it may be sensitive to the starting location and the order in which each link is checked. I mitigate this by utilizing various sequences and obtaining the counterfactual prices that lie in a range.

In this study, a structural change like the Arch-Triton merger enters into the model through two aspects. Firstly, ownership changes set foot on the companies' profit maximization problems through Nash bargaining and the threat of replacement. Secondly, Arch assimilated North Rochelle as part of Black Thunder after the merger, which resulted in fewer varieties and distinct production capacities. Black Thunder grasps more production capacity after razing the boundary in between and incorporating it with North Rochelle.

Appendix B.3 contains the algorithms I use for the counterfactual. Roughly speaking, it is an extension of Fan and Yang (2020) to accustom the capacity constraints and the threat of replacement. Analogous to Algorithm 1, I define the counterfactual algorithm as a nested fixed-point problem and decompose it into an inner and outer loop. In the inner loop, for the network T at hand, I calculate the associated best response prices for all the inside goods, with the no-replacement conditions satisfied. For this model, The best response price $p_{ijt} = \min\{p_{ijt}^{Nash}, p_{ijt}^{NR}\}$. p_{ijt} constitutes the price vector P associated with network T. P is an interconnected system and solved by a fixed-point iteration in Ho and Lee (2019) that iterating over $\forall j \in T$ until convergence. Sometimes I may have an unstable equilibrium, where I use the derivative-free spectral algorithm initialized in Barzilai and Borwein (1988) to improve the convergence rate.

In the outer loop, I iterate over the set of products ensuing a pre-determined order and check each unilateral deviation of coal companies for good j. The single deviation refers to unilateral addition and cut-off. While unilateral has a direct interpretation that coal company considering selling j to a potential customer i', cut-off embodies two scenarios. If current p_{ijt} is smaller than the marginal cost, coal companies may shut off the link without discussing it with the plants.

Also, coal companies may find that, given the current network T, instead of selling to i, selling to $i' \in \mathcal{H}$ is more profitable, so they switch to serving a different set of customers. This "replacement" is different from the p_{ijt}^{NR} in the inner loop. In the inner loop, the p_{ijt}^{NR} refer to the price that plant

64. Whether or not this is a drawback is ambiguous. After a structure change, I may expect the market to restore its stability by shifting into the stable network closest to the current structure.

i would not want to replace j with j', which does not promise that the coal companies would not want to deviate from selling to i. This extra flexibility speeds up the convergence[65]. The outer loop terminates if and only if no more coal company wants to deviate from the current network unilaterally.

When simulating the counterfactual, I assume the long-term contract follows an exogenous process. The decisions regarding spot purchases are conditional on the current long-term contract. Similarly, I presume that power plants are flexible in substituting the purchase from contracts or spot market at the cost of penalty as in (22).

The paragraphs below discuss two counterfactuals. Firstly, what if there is no merger at all? Secondly, what if there is no divestiture of Buckskin after the merger? Since the fact is the successful consolidation of Black Thunder and North Rochelle, the counterfactual would keep the original market structure. Without that Arch-Triton merger, North Rochelle remains a separate coal mine, combined with Buckskin, owned by Triton.

Furthermore, recall that a divestiture remedy of Buckskin mine to a previously fringe coal producer Kiewit was proposed to alleviate the anti-competitive effect. While denied by the FTC, the court accepted this merger remedy. FTC argued that the Arch-Triton merger would trigger entire price collusion among the major PRB producers. This counterfactual evaluates the pro-competitive power of this divestiture when there is no tacit collusion, price coordination, or other anti-competitive behavior out of the scope of this study. Theoretically, without divestiture, besides the direct effect from the joint optimization after the ownership change, Buckskin, now a part of Arch, would no longer be a threat when bargaining with coals from Arch. This would drive prices up even without the hazard of potential price coordination suspected by the FTC.

In addition, suppliers' capacity constraints also enter as an important factor in determining the new equilibrium prices. Structural changes like merger, acquisition, entry, or exit are always accompanied by generation, dissolution, and transfers of capacities. Capacity constraints are directly connected with prices or margins, as discussed in Pakes (2010) for the health sector. Wang et al. (2020) in a reduced-form study shows that coal prices in China increase with the de-capacity policy. In a static setting, the influence becomes stronger since a coal company wants to sell up

65. This replacement is indeed another version of the no-addition condition in a stable network. So it does not contradict the conditions in supply-side estimation.

Source	Without Merger
(i) Direct ownership change	Price decreases for the plants currently engaged in transactions with merged entities.
(ii) Indirect ownership change	Price decreases for the plants that previously using merged entities as the binding threat. Otherwise there would be no influence. Has the possibility to make the current network unstable.
(iii) Cost efficiency	Price decreases, but I do not allow this due to the parsimonious marginal cost in (21).
(iv) Change in capacity constraints	It would always invalidate the current network and set off iterations toward a new stable network. The influence on the prices is undetermined since it depends on the step-by-step deviation to form the new stable network.

Table 8: Summary for different factors that would change price in the model.

to its production in each period. Higher capacities transform into large storage in the site, waiting for sales, which may lead to a less aggressive pricing behavior of suppliers, especially when the consumers have heterogeneity in both fixed costs and tastes. Moreover, changes in productive capacities would deprive the current network's stability and set about the paths toward a new equilibrium. Unfortunately, no theory can predict the direction of spot prices when the wheel stops.

Table 8 briefly summarizes the possible forces that would influence the ultra spot prices without the Arch-Triton merger, while Table 9 illustrates in detail what one may expect for the counterfactuals when some of the model primitives are shut down. Due to the lack of ability to flexibly estimate the marginal cost, though efficiency is the major justification for Arch-Triton merger, currently I ignore the possible change in marginal cost after the merger.

Figure 15 reports the counterfactual results about the spot price change from setting 1, where the capacity constraint equals to the observed traded volumes. Figure 16 shows the corresponding consumer surplus (CS) in each counterfactual settings. Both figures contains four different cases, summarised below[66].

Case 1 – No merger. North Rochelle's capacity comes equally from all other coals.

Case 2 – No merger. North Rochelle's capacity all comes from Black Thunder.

Case 3 – No merger. Extra capacity for North Rochelle.

Case 4 – Merge but no divestiture. All else is the same.

66. Recall that after merge, North Rochelle is embodied into Black Thunder so that I do not observe its trading volumes separately. To impute, I take the average of North Rochelle's previous sales.

In Cases 1 and 2, the network before the structure change is not feasible anymore, while Cases 3 and 4 regard the original network as feasible but may not be profitable. Case 3 has a capacity expansion, where the original network is feasible regarding capacity. However, this capacity expansion, combined with the threat of replacement, has the possibility to drive up the price instead. The reason behind this would be that current low prices usually are dragged down by the threat of replacement. And with the network expansion, the threat of replacement drops to marginal cost, which may cause the links with the lowest prices to be cut off at the first few iterations. Case 4, on the other hand, only has ownership change, but with capacities staying the same.

The vertical lines in Figure 15 and 16 refer to multiple stable networks from the same starting point through different orders of $j \in \mathbb{J}_{it}$ and $i \in \mathbb{I}^d$ in Algorithm 3 [67]. Under the binding capacity constraint in setting 1, most counterfactual prices are widespread. The randomness in regaining the network stability dominates the ownership change and the incorporation of North Rochelle into Black Thunder. Prices change become very specific to the model settings. For instance, FOB prices may decrease without the merger in 2006Q3, but the decrease is no more guaranteed after capacity restrictions complicate the system. Despite the set results from the counterfactual, it seems that divestiture indeed lowers the FOB prices.

Though the capacity constraints are exogenous and fixed, how many quantities are sold in time t is a summation of the endogenous choice inside the model. After a structural change, with a prohibitively high punishment for violation in capacity restrictions, a stable network that fulfill all the accessible may not exist[68]. This leads to a decrease in consumer surplus and proportion of capacities traded, as shown in Figure 16 regardless of the FOB prices. Alternatively, as shown in the bottom panel of Figure 16, consumer surplus suffers more from a decrease in trading volumes rather than the FOB prices.

A supplementary Figure 17 illustrates how a new stable network is reached from the starting value. For this specific example[69], the network takes around 120 iterations to regain its stability. Panel (a) in Figure 17 records each step's averaged FOB prices and transportation costs. FOB prices decrease into unreasonably negative value at first, an indicator of a non-stable network since sup-

67. It is clear that a new stable network tends to exist, though it is definitely not unique. I only attempt six to seven different paths, some of which may not converge. By no means would those paths exhaust all the possibilities.

68. In Algorithm 3, coal companies optimize over a block-wise greedy algorithm, which is known to have spared room usually. The initial network uses up all the capacities by definition.

69. This example comes from one calculation from Case 3, 2006Q3.

pliers would want to cut those links. Then prices readily go back until convergence is obtained at around \$19 to \$20, slightly below the initial value. The average transportation costs are slightly below the original value as well. Panel (b) of Figure 17 shows the number of realized transactions or solid links. The number of solid links becomes the same when the network regains stability, yet around 20% of the links in the original network are replaced.

8 Conclusion

This study develops a comprehensive model depicting transactions among individuals that incorporates two crucial factors – endogenous network formation and capacity constraints. I discuss its estimation and computational algorithms. Then I apply this model to a real-world example where a debatable merger has taken place and demonstrate that all those features in the model matter in the analysis.

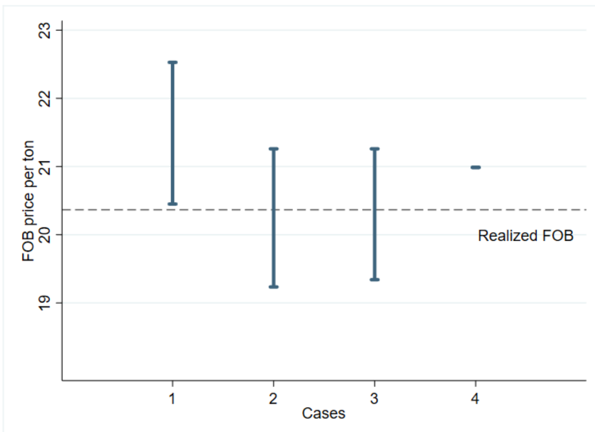
This research has several drawbacks. First, Algorithms 1 and 3 are still highly computationally intensive. The model I solved in this study is at the edge of infeasible [70], which limits the scope of other applications and extensions. At the current stage, the best usage may immerse in low dimensional cases. Besides, I make a bunch of simplifications, which usually lack real-world grounds. For instance, computational difficulty drives a parsimonious parametrization of marginal cost in (21), which fails to account for the cost synergies emphasized by Arch to justify this merger.

Besides finding a better specification to improve computation, further research may also go in the direction of the way to deal with the co-existence of spot market and long-term contracts. By shutting down the capacity constraint, adding dynamics becomes achievable even when there is endogenous network formation. Such a comprehensive model may consummate the study of the coal-plant relationship. Modeling how the current spot prices would be incorporated into long-term contracts is another aspect of understanding this market better. In this study, I also pull away from the three-party structure of coal delivery. It may be interesting to have the transportation agents enter as endogenous participants and evaluate their reaction after a structure change among coal companies.

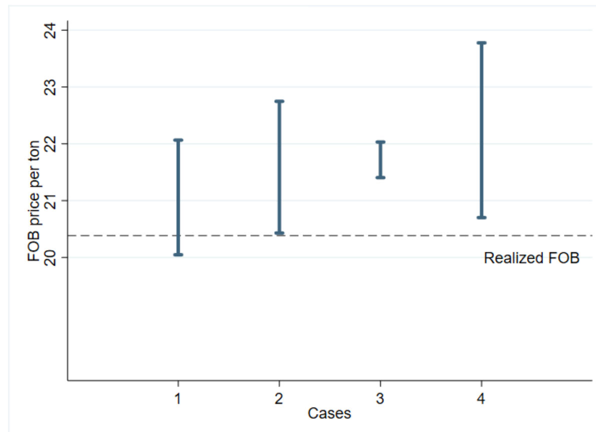
70. Here is the configuration of the computer I use for this study: AMD Ryzen 9 3900X 12-Core Processor, 3.79 GHz, 64RAM. If possible, all the computational procedure is paralleled. Even so, it takes more than twenty days for Algorithm 1 to go over all the discrete points in the grid search. Moreover, it takes one day to a week for Algorithm 3 to reach a stable network.

	No Network Change	Network Change
No Capacity Constraint	Prices would decrease with no merger and would increase with no divestiture. The structure change influences prices through the ownership matrix when optimization. With no merger, there is an increase in consumers' welfare and vice versa for no divestiture.	Both plants and coal companies can deviate from the current network when stable prices decrease or increase. In the new situation, the network before the structure change is feasible but not favorable. Without capacity constraints, plants can always deviate and get coals at their marginal cost. In the beginning, for instance, with no divestiture, p_{ijt} may increase if it is generated by Nash bargaining or threat of replacement from the merged entities. However, even when network formation is allowed, if we have a unified marginal cost, there may not be any network change.
Capacity Constraint	A structural change usually is accompanied by a change in capacities. Two effects are present. Ownership changes have a direct effect on prices from FOC. And different capacities may change the credible replacement threats plants have. For instance, if there is no divestiture, Buckskin's capacity adds to Arch's products, which makes Buckskin no longer a valid threat when bargaining with Arch. Both effects move in the same direction, reducing prices with no merger and increasing with no divestiture.	The initial network becomes infeasible due to the violation of capacity constraints, which forces a change in the network even with a unified marginal cost. Coal companies' sales choices become interconnected. One may end up in a conspicuously different network with an unknown direction of prices after the structure change. All the participants in the network may be influenced.

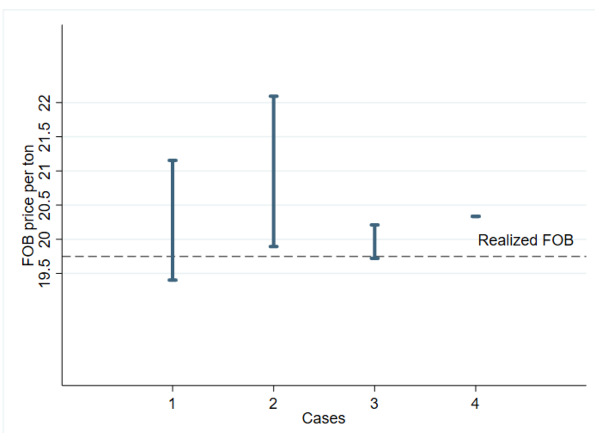
Table 9: Possible counterfactuals from different model primitives.



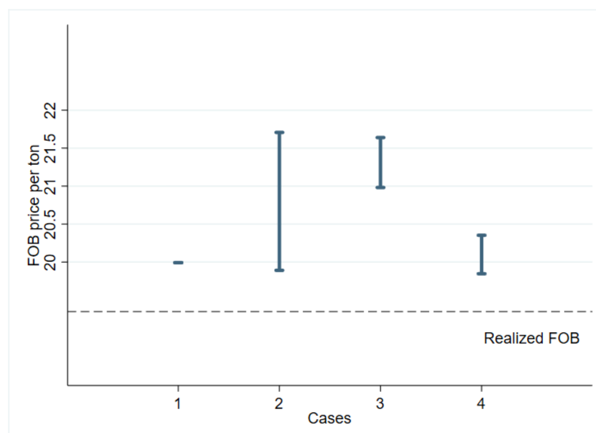
(a) 2006Q3



(b) 2006Q4

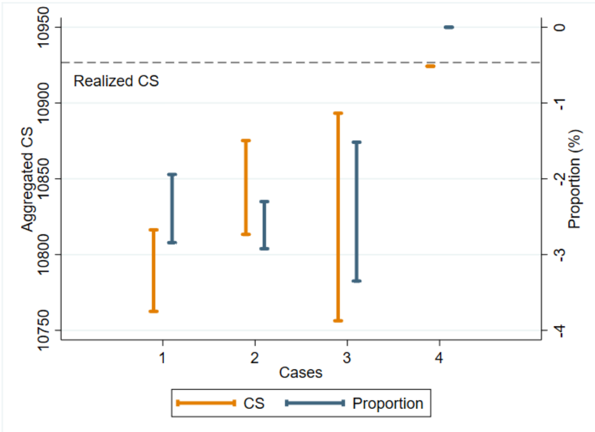


(c) 2007Q1

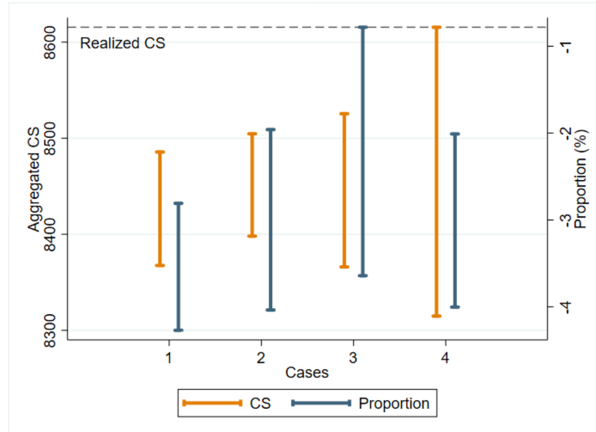


(d) 2007Q2

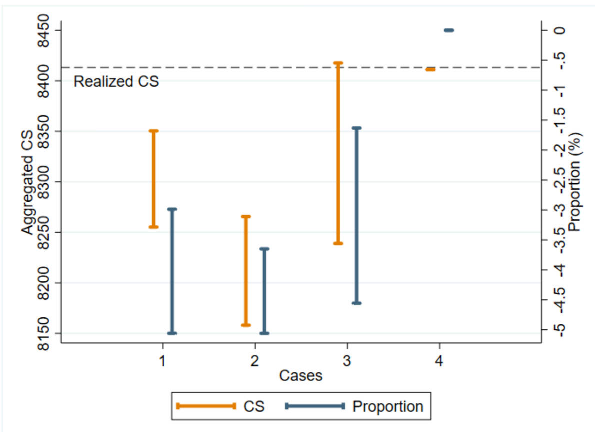
Figure 15: Counterfactual – the average of FOB prices.



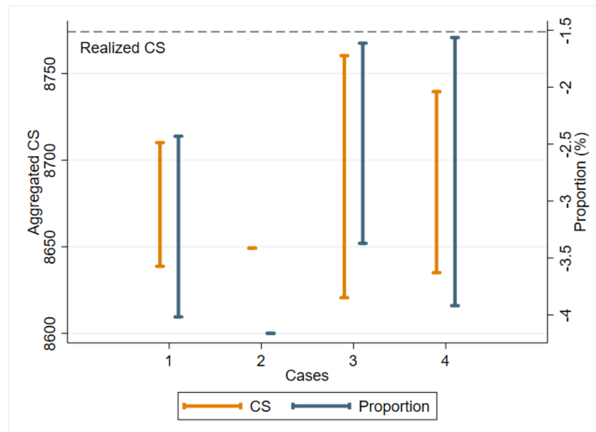
(a) 2006Q3



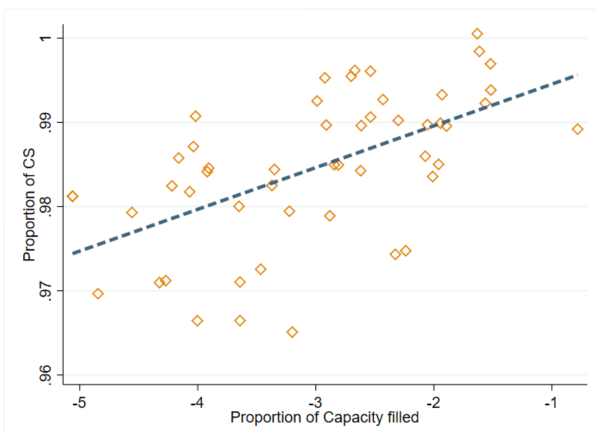
(b) 2006Q4



(c) 2007Q1

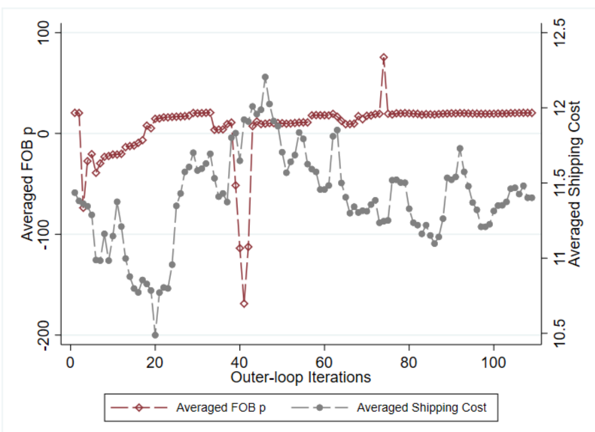


(d) 2007Q2

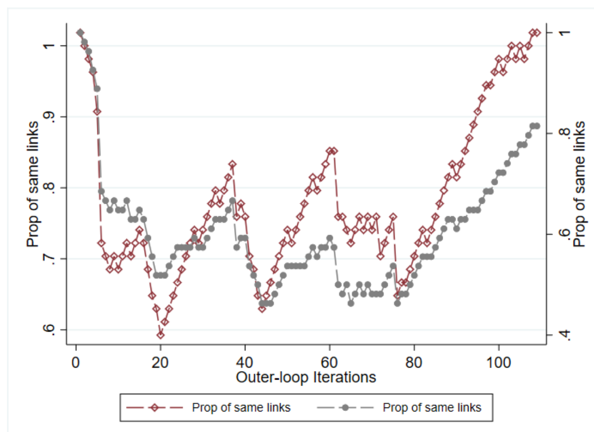


(e) Relation between CS and capacity fulfills

Figure 16: Counterfactual – the consumer surplus and the fulfillment of capacity.



(a) Averaged prices and transportation costs for the purchased goods in the current iteration



(b) Proportion of solid links in the current iteration

Figure 17: The paths towards a new stable network, in a randomly-chosen case.

	Western SPRB	Central Appalachia	Northern Appalachia	Mid-Continent Illinois Basin	Western Cen. Rockies	Mid-Continent Gulf Lignite	Western Four Corners	Western NPRB
Product Characteristics								
$/MMBtu FOB	0.3182	1.0688	0.8168	0.8549	0.7151	0.8964	0.9395	0.4160
	(0.1590)	(0.4570)	(0.4656)	(0.3509)	(0.3567)	(0.4515)	(0.3793)	(0.2410)
$/MMBtu ALL	0.9586	1.4404	1.0281	1.0759	1.1493	1.0218	1.1601	1.0681
	(0.3470)	(0.5449)	(0.5896)	(0.4319)	(0.5160)	(0.4986)	(0.4168)	(0.4069)
MMBtu/ton	17.3513	24.8291	24.0501	22.5693	23.0313	12.8643	19.6138	18.3574
	(0.5123)	(1.1442)	(3.4284)	(1.5288)	(1.7251)	(1.4634)	(1.4024)	(0.8576)
SO2MMBtu	0.7091	1.5655	3.3559	4.3680	0.9437	3.3169	1.4011	0.9988
	(0.2817)	(0.6339)	(1.6051)	(1.9065)	(0.3827)	(1.7488)	(0.5013)	(0.4381)
Ash %	5.1200	10.7617	13.4482	9.7209	9.5232	16.5267	16.2789	5.8215
	(0.7727)	(2.8116)	(8.7631)	(3.2982)	(3.8694)	(4.2779)	(4.0814)	(2.2547)
Tons(000)	102.5264	26.7577	30.5664	42.7190	56.4536	298.4040	257.6731	109.2597
	(113.4298)	(32.0885)	(51.8146)	(55.8972)	(61.3089)	(216.8247)	(249.7318)	(113.0711)
Proportion of Deliveries								
# Obs	34992	66297	32209	20957	8726	1647	1500	3331
Long-term	0.78	0.62	0.66	0.69	0.78	0.97	0.89	0.85
Spot	0.21	0.37	0.31	0.29	0.22	0.03	0.11	0.15
Unknown	0.01	0.00	0.03	0.02	0.00	0.00	0.00	0.00

Standard errors in parentheses. Prices are measured in 1998 dollars.
Original data is in monthly frequency. An observation is the monthly delivery of coals from a coal mine to a power plant.

Table 10: Summary of Characteristics of Coal Delivered, by Mine Supply Region: 1998-2008. The table only contains the eight largest coal supply regions.

	Western SPRB	Central Appalachia	N. Appalachia Northeast	Mid-Continent Illinois Basin	Western Cen. Rockies	Mid-Continent Gulf Lignite	N. Appalachia Ohio	Western NPRB
Product Characteristics								
$/MMBtu FOB	0.3284	1.1606	1.0116	0.9890	0.8093	0.9985	0.9346	0.4240
	(0.1496)	(0.3605)	(0.2526)	(0.2856)	(0.3228)	(0.2927)	(0.3228)	(0.2255)
$/MMBtu ALL	0.9771	1.5395	1.2858	1.2866	1.4284	1.1187	1.1468	1.1122
	(0.3071)	(0.4072)	(0.3589)	(0.3360)	(0.3893)	(0.3498)	(0.3598)	(0.3555)
MMBtu/ton	17.3545	24.6856	25.5791	22.4483	23.3860	13.0473	23.8292	18.3734
	(0.5264)	(1.1893)	(1.0264)	(1.9661)	(1.3192)	(0.9408)	(1.2394)	(0.7259)
$SO2/MMBtu	0.1403	0.2557	0.5895	0.7179	0.1845	0.4632	0.7869	0.1788
	(0.1287)	(0.2664)	(0.5589)	(0.7539)	(0.1850)	(0.4202)	(0.7211)	(0.1689)
Ash %	5.1152	11.0727	9.1245	9.3966	8.9429	15.7748	10.6797	5.6852
	(0.6899)	(2.9661)	(2.4909)	(3.5133)	(2.0647)	(2.8499)	(2.3554)	(2.1517)
Tons(000)	255.5041	52.2530	74.9967	78.0611	111.2616	1003.7130	140.9790	274.7213
	(351.6465)	(69.8035)	(102.2003)	(109.3504)	(138.2851)	(262.0504)	(329.2238)	(114.2103)
Proportion of Deliveries								
# Obs	14336	9206	3826	2752	1687	351	1099	1305
Long-term	0.77	0.53	0.67	0.62	0.70	0.98	0.53	0.83
Spot	0.23	0.47	0.33	0.38	0.30	0.02	0.47	0.17

Standard errors in parentheses. Prices are measured in 1998 dollars.
Original data is in monthly frequency. An observation is the monthly delivery of coals from a coal mine to a power plant.
Here, $SO2/MMBtu is the sulfur dioxide allowance needs to pay at the average level per MMBtu.
It is what actually used in the estimation, different from what the SO2MMBtu represents in the summary table for the original data.
The small fraction of data containing unknown contract type is dropped.
Purchased quantities are large because it is an aggregation from month to quarter.

Table 11: Summary of Characteristics of Coal Delivered to the Power Plants used in the Structural Model, by Mine Supply Region: 1998-2008. The table only contains the eight largest coal supply regions.

References

Adao, R., A. Costinot, and D. Donaldson. 2017. "Nonparametric Counterfactual Predictions in Neoclassical Models of International Trade." *American Economic Review* 107 (3): 633–689.

Aguirregabiria, V., and M. Marcoux. 2021. "Imposing equilibrium restrictions in the estimation of dynamic discrete games." *Quantitative Economics* 12 (4): 1223–1271.

Baker, J., and T. Bresnahan. 1985. "The Gains from Merger or Collusion in Product-Differentiated Industries." *The Journal of Industrial Economics* 33 (4): 427–444.

Barzilai, J., and J. Borwein. 1988. "Two-Point Step Size Gradient Methods." *IMA Journal of Numerical Analysis* 8 (1): 141–148.

Berry, S. 1994. "Estimating Discrete-Choice Models of Product Differentiation." *The RAND Journal of Economics* 24 (2): 242–262.

Berry, S., J. Levinsohn, and A. Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–890.

Bhat, C. 2005. "A Multiple Discrete-Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions." *Transportation Research Part B* 39 (8): 679–707.

———. 2008. "The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions." *Transportation Research Part B* 42 (3): 274–303.

Blundell, R., and R. Matzkin. 2014. "Control Functions in Nonseparable Simultaneous Equations Models." *Quantitative Economics* 5 (2): 271–295.

Bontemps, C., C. Gualdani, and K. Remmy. 2022. "Price Competition and Endogenous Product Choice in Networks: Evidence from the US Airline Industry." Working Paper, https://events.barcelonagse.eu/live/files/3773-bontempsgualdaniremmy2022pdf, January.

Chandra, A., and M. Weinberg. 2018. "How Does Advertising Depend on Competition? Evidence from U.S. Brewing." *Management Science* 64 (11): 4967–5460.

Cicala, S. 2015. "When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation." *American Economic Review* 105 (1): 411–444.

Crawford, G., and A. Yurukoglu. 2012. "The Welfare Effects of Bundling in Multichannel Television Markets." *American Economic Review* 102 (2): 643–85.

Davis, L., and C. Hausman. 2016. "Market Impacts of a Nuclear Power Plant Closure." *American Economic Journal: Applied Economics* 8 (2): 92–122.

Davis, L., and L. Kilian. 2011. "The Allocative Cost of Price Ceilings in the U.S. Residential Market for Natural Gas." *Journal of Political Economy* 119 (2): 212–241.

Dedieu, A., H. Hazimeh, and R. Mazumder. 2021. "Learning Sparse Classifiers: Continuous and Mixed Integer Optimization Perspectives." Working Paper, https://arxiv.org/pdf/2001.06471.pdf.

Dehejia, R., and S. Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs." *Journal of the American Statistical Association* 94:1053–1062.

Doyle Jr., R., and N. Ray. 2005. "Role of Customer Complaints and Testimony in Merger Enforcement Post-Arch Coal and Oracle." *The Sedona Conference Journal* 6.

Dubé, J., A. Hortaçsu, and J. Joo. 2021. "Random-Coefficients Logit Demand Estimation with Zero-Valued Market Shares." *Marketing Science* 40 (4): 637–660.

Fan, Y. 2013. "Ownership Consolidation and Product Characteristics: A Study of the US Daily Newspaper Market." *American Economic Review* 103 (5): 1598–1628.

Fan, Y., and C. Yang. 2020. "Competition, Product Proliferation, and Welfare: A Study of the US Smartphone Market." *American Economic Journal: Microeconomics* 12 (2): 99–134.

Friberg, R., and A. Romahn. 2015. "Divestiture requirements as a tool for competition policy: A case from the Swedish beer market." *International Journal of Industrial Organization* 42:1–18.

Gandhi, A., and J. Houde. 2019. "Measuring Substitution Patterns in Differentiated Products Industries." Working Paper,https://jfhoude.wiscweb.wisc.edu/wp-content/uploads/sites/769/2018/08/GH_v6.pdf.

Ghili, S. 2022. "Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance." *Marketing Science* 41 (3): 501–527.

Gowrisankaran, G., A. Nevo, and R. Town. 2015. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review* 105 (4): 173–203.

Grennan, M. 2013. "Price Discrimination and Bargaining: Empirical Evidence from Medical Devices." *American Economic Review* 103 (1): 145–177.

Gross, A. 2019. "Private Labels and Bargaining in the Supply Chain." Working Paper,https://drive.google.com/file/d/1maC-RWHQca0nAsEKCt9VXHa3bJpaTqC2/view.

Heckman, J., H. Ichimura, and P. Todd. 1997. "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme." *The Review of Economic Studies* 64 (4): 605–654.

Ho, K., and R. Lee. 2019. "Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets." *American Economic Review* 109 (2): 473–522.

Hong, M., M. Razaviyayn, Z. Luo, and J. Pang. 2016. "A Unified Algorithmic Framework for Block-Structured Optimization Involving Big Data: With applications in machine learning and signal processing." *IEEE Signal Processing Magazine* 33 (1): 57–77.

Houde, J. 2012. "Spatial Differentiation and Vertical Mergers in Retail Markets for Gasoline." *American Economic Review* 102 (5): 2147–2182.

Jha, A. 2019. "Dynamic Regulatory Distortions: Coal Procurement at U.S. Power Plants." Working Paper. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3330740.

Kaplan, S. 2007. "Rail Transportation of Coal to Power Plants: Reliability Issues." Congress Report.

Kim, T., and J. Villas-Boas. 2018. "A Note on Demand Estimation with Supply Information in Non-Linear Models." Working Paper, http://faculty.haas.berkeley.edu/VILLAS/2.pdf.

Kline, B., A. Pakes, and E. Tamer. 2021. "Moment Inequalities and Partial Identification in Industrial Organization." Working Paper, https://www.nber.org/system/files/working_papers/w29409/w29409.pdf, October.

La Cruz, W., J. Martinez, and M. Raydan. 2006. "Spectral Residual Method Without Gradient Information for Solving Large-scale Nonlinear System of Equations." *Mathematics of Computation* 75 (255): 1429–1448.

Lee, J., and K. Seo. 2015. "A Computationally Fast Estimator for Random Coefficients Logit Demand Models using Aggregate Data." *RAND Journal of Economics* 46 (1): 86–102.

Liebman, E. 2018. "Bargaining in Markets with Exclusion: An Analysis of Health Insurance Networks." Working Paper, https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxlbGlibGllYm1hbnxneDoxMzdlNjg2ZWRhMWVlNzM3.

Miller, N., G. Sheu, and M. Weinberg. 2021. "Oligopolistic Price Leadership and Mergers: The United States Beer Industry." Working Paper, http://www.nathanhmiller.org/priceleadership.pdf.

Miller, N., and M. Weinberg. 2017. "Understanding the Price Effects of the MillerCoors Joint Venture." *Econometrica* 85 (6): 1763–1791.

Nevo, A. 2000. "Mergers with Differentiated Products: The Case of the Ready-to-Eat Cereal Industry." *RAND Journal of Economics* 31 (3): 395–421.

Pakes, A. 2010. "Alternative Models for Moment Inequalities." *Econometrica* 78 (6): 1783–1822.

Sheu, G. 2014. "Price, Quality, and Variety: Measuring the Gains from Trade in Differentiated Products." *American Economic Jounrnal: Applied Economics* 6 (4): 66–89.

Tenn, S., L. Froebb, and S. Tschantz. 2010. "Mergers when firms compete by choosing both price and promotion." *International Journal of Industrial Organization* 28 (6): 695–707.

Tenn, S., and J. Yun. 2011. "The success of divestitures in merger enforcement: Evidence from the J&J–Pfizer transaction." *International Journal of Industrial Organization* 29 (2): 273–282.

Train, K., and W. Wilson. 2011. "Coal Demand and Transportation in the Ohio River Basin: Estimation of a Continuous/Discrete Demand System with Numerous Alternatives." Working Paper, https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.491.4959&rep=rep1&type=pdf.

Valero, A., A. Valero, G. Calvo, and A. Ortego. 2018. "Material Bottlenecks in the Future Development of Green Technologies." *Renewable and Sustainable Energy Reviews* 93:178–200.

Wang, X., C. Liu, S. Chen, L. Chen, K. Li, and N. Liu. 2020. "Impact of Coal Sector's De-Capacity Policy on Coal Price." *Applied Energy* 265.

Wang, X., R. Mittelhammer, T. Marsh, and J. McCluskey. 2017. “Divestiture of US Business May Fail as a Merger Remedy: The Case of the US Beer Industry.” Working Paper, https://ideas.repec.org/p/ags/aaea17/258388.html.

Zahur, N. 2022. “Long-term Contracts and Efficiency in the Liquefied Natural Gas Industry.” Working Paper, https://uc513cacf3a5b5deb7d019bf85c1.dl.dropboxusercontent.com/cd/0/inline2/BwNqdJ4gXZXykTnyHAfdbGWqlBHwq_yYI7aJgVHolkLcKgWsqe2uKD426otKa8ynvdecmhhL9sdzV6wmUFfe7ay0ClreJgTTM5qQRSKqaGRhx_WoOex0nnmnGK1-waBs6wGh1_3aYrrjwcDj8gnz2eyyczpPpee_LKXUhPadfS09LLb16GQJahV5rLcySnwHwHftDqnbSMwH1kOva29vqIBCm39rh68LQJIqwzLs3cUa2N414ke8tRmuW3mSsEVeNpe1q7ESL9N59BvmtvQ2G_UuZmGxz19Ov0hc0VyUSVtvQ9Wni6k9xw0ncO-73AHkytf93x3IwsRyiC42xEId3jwTZTeS41MXsXmInCVw_Apy4Ccb2Iz4bIIW_FUNPK_KlAJ49G6IsTSVmngRCLMmhmHC9MQaPJlulRaHPpVhsLfcog/file, September.

Zheng, M., J. Lin, X.-M. Yuan, and E. Pan. 2019. “Impact of an Emergency Order Opportunity on Supply Chain Coordination.” *International Journal of Production Research* 57 (11): 3504–3521.

Appendix to

Bargaining, Merger, Capacities, and Endogenous Network Formation: the Case of Power Plants and Coal Suppliers in the US

Appendix A Derivations omitted in the Main Text

Appendix A.1 Derivation of Nested CES

The objective function for plant i at t is to minimize the overall static cost of coal procurement at the current period t.

$$\min_{Q_{ijt}} C_{it} = \left(\sum_{g \in Gen_{it}} \left[\left(\sum_{j=1}^{J_g} \phi_{ijt}^{1-\rho} Q_{ijt}^{\rho} \right)^{\frac{1}{\rho}} \right]^{\sigma} \right)^{\frac{1}{\sigma}}, \quad \text{s.t.} \sum_{j=1}^{J_{it}} h_{ijt} Q_{ijt} \geqslant E_{it} \tag{A.1}$$

Or in the expression of market share, rather than overall quantities,

$$\min_{S_{ijt}} C_{it} = \left(\sum_{g \in G_{it}} \left[\left(\sum_{j=1}^{J_g} \phi_{ijt}^{1-\rho} \left(\frac{S_{ijt}}{h_{ijt}} \right)^{\rho} \right)^{\frac{1}{\rho}} \right]^{\sigma} \right)^{\frac{1}{\sigma}}, \quad \text{s.t.} \sum_{j=1}^{J_{it}} S_{ijt} = 1 \tag{A.2}$$

This can be derived by $Q_{ijt} = (E_{it} \times S_{ijt})/h_{ijt}$, and S_{ijt} is the energy share. One can take out E_{it}, since E_{it} is the same for all purchased j. Moreover, E_{it} is determined outside the nested CES, the optimal solution for energy share S_{ijt} is obtained given E_{it} and does not depend on it. Sheu (2014) build up a direct connection between nested CES and nested Logit, with the nested CES interpreted as a modified Logit. Therefore, the estimation of nested CES with endogenous prices follows directly from Berry (1994), where one can decompose the observed unconditional market share by $S_{ijt} = S_{ijt|igt} \times S_{igt}$. Among the $J_{g_{it}}$ goods in nest g_{it}, plant i solves problem A.3 to determine $S_{ijt|igt}$,

$$\min_{Q_{ijt}, j \in J_{g_{it}}} \sum_{j=1}^{J_{g_{it}}} \phi_{ijt}^{1-\rho} \left(\frac{S_{ijt|igt}}{h_{ijt}} \right)^{\rho}, \quad \text{s.t.} \sum_{j=1}^{J_{g_{it}}} S_{ijt|igt} = 1 \tag{A.3}$$

The Lagrange is $\mathscr{L} = \sum_{j=1}^{J_g} \phi_j^{1-\rho} \left(\frac{S_{ijt}}{h_{ijt}} \right)^{\rho} + \lambda_1 (S_{igt} - \sum_{j=1}^{J_g} S_{ijt})$ implies the FOC,

$$S_{ijt} = \left(\frac{\lambda_1 h_{ijt}^{\rho}}{\rho} \right)^{\frac{1}{\rho-1}} \phi_{ijt}, \quad \text{and} \quad \lambda_1 = \left[\frac{S_{igt}}{\sum_{m=1}^{J_g} (\frac{h_{imt}^{\rho}}{\rho})^{\frac{1}{\rho-1}} \phi_{imt}} \right]^{\rho-1} \tag{A.4}$$

Then insert (A.4) into the binding budget constraint in (A.3), $S_{ijt|igt} = (\phi_{ijt} h_{ijt}^{\frac{\rho}{\rho-1}}) / (\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}})$, which has a similar structure as the nested Logit, except for the h_{ijt}-adjusted baseline cost.

Upper layer determines S_{igt} with the optimal solution $S_{ijt|igt}$.

$$\min_{S_{igt}, g_{it} \in G_{it}} \sum_{g_{it} \in G_{it}} M_{g_{it}}^{\sigma}, \quad \text{s.t.} \quad M_{g_{it}} = \left[\sum_{j=1}^{J_g} \phi_{ijt}^{1-\rho} \left(\frac{S_{ijt|igt} \times S_{igt}}{h_{ijt}} \right)^{\rho} \right]^{\frac{1}{\rho}}, \sum_{g_{it} \in G_{it}} S_{igt} = 1 \tag{A.5}$$

Insert the lower-level $S_{ijt|igt}$ into (A.5),

$$M_{g_{it}}^{\sigma} = \left[\frac{1}{\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}}} \right]^{\sigma} \times \left[\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right]^{\frac{\sigma}{\rho}} \times S_{igt}^{\sigma} \tag{A.6}$$

Lagrange $\mathcal{L} = \sum_{g_{it} \in G_{it}} M_{g_{it}}^{\sigma} + \lambda_2 (1 - \sum_{g_{it} \in G_{it}} S_{igt})$ and further calculation gives,

$$S_{igt} = \frac{\left(\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\frac{\sigma(\rho-1)}{\rho(\sigma-1)}}}{\sum_{g_{it} \in G_{it}} \left(\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\frac{\sigma(\rho-1)}{\rho(\sigma-1)}}} \tag{A.7}$$

Therefore, the unconditional energy share can be expressed as

$$S_{ijt} = S_{ijt|igt} \times S_{igt} = \frac{\phi_{ijt} h_{ijt}^{\frac{\rho}{\rho-1}}}{\left(\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\left[1 - \frac{\sigma(\rho-1)}{\rho(\sigma-1)}\right]} \times \sum_{g_{it} \in G_{it}} \left(\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}} \right)^{\frac{\sigma(\rho-1)}{\rho(\sigma-1)}}} \tag{A.8}$$

The price elasticities implied by this CES demand system, with level prices entering into the cost function, follows the equations below,

$$Ed_{ijt} = \frac{\partial S_{ijt}}{\partial p_{ijt}} \frac{p_{ijt}}{S_{ijt}} = \beta_p p_{ijt} \times \left[1 - \left(\left[1 - \frac{\sigma(\rho-1)}{(\sigma-1)\rho} \right] S_{ijt|igt} + \frac{\sigma(\rho-1)}{(\sigma-1)\rho} S_{ijt} \right) \right]$$

$$Ed_{ijkt} = \frac{\partial S_{ijt}}{\partial p_{ikt}} \frac{p_{ikt}}{S_{ijt}} = \begin{cases} -\dfrac{S_{ikt}}{S_{ijt}} \beta_p p_{ikt} \times \left(\left[1 - \dfrac{\sigma(\rho-1)}{(\sigma-1)\rho} \right] S_{ijt|igt} + \dfrac{\sigma(\rho-1)}{(\sigma-1)\rho} S_{ijt} \right), \; j, k \in J_{igt} \\ -S_{ikt} \beta_p p_{ikt} \times \dfrac{\sigma(\rho-1)}{(\sigma-1)\rho}, \quad j \in J_{igt}, k \notin J_{igt} \end{cases}$$

The cross-price elasticities varies with whether or not j and k are in the same nest. Nested CES expects more substitute inside rather than across a nest. Therefore a correctly-specified nest structure needs $\hat{\sigma} > \hat{\rho} > 1$.

The equation (8) in the main text is derived below. If $j, j' \in J_{igt}$, then the (8) follows directly, since $\left(\sum_{m \in J_{g_{it}}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}}\right)^{\left[1-\frac{\sigma(\rho-1)}{\rho(\sigma-1)}\right]}$ from both j and j' cancel out. If $j \in J_{igt}, j' \notin J_{igt}$, then

$$
\begin{aligned}
\ln\left(\frac{S_{ijt}}{S_{ij't}}\right) &= \ln\left(\frac{\phi_{ijt}}{\phi_{ij't}}\right) + \frac{\rho}{\rho-1}\ln\left(\frac{h_{ijt}}{h_{ij't}}\right) - \left(1 - \frac{\sigma-1}{\sigma}\frac{\rho}{\rho-1}\right)\left[\ln \frac{\sum_{m \in J_{igt}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}}}{\sum_{m \in J_{ig't}} \phi_{imt} h_{imt}^{\frac{\rho}{\rho-1}}}\right] \\
&= \ln\left(\frac{\phi_{ijt}}{\phi_{ij't}}\right) + \frac{\rho}{\rho-1}\ln\left(\frac{h_{ijt}}{h_{ij't}}\right) - \left(1 - \frac{\sigma-1}{\sigma}\frac{\rho}{\rho-1}\right)\frac{\ln(S_{igt}) - \ln(S_{ig't})}{\frac{\sigma}{\sigma-1}\frac{\rho-1}{\rho}} \\
&= (W_{ijt} - W_{ij't})\beta_X + (P_{ijt} - P_{ij't})\beta_p + \xi_{ijt} - \xi_{ij't} + \frac{\rho}{\rho-1}[\ln(h_{ijt}) - \ln(h_{ij't})] \\
&\quad + \left(1 - \frac{\sigma-1}{\sigma}\frac{\rho}{\rho-1}\right)\Big[log(S_{igt}) - log(S_{ig't})\Big]
\end{aligned}
$$

Further simplification as in Berry (1994) is impracticable here, since the denominator j' is not the outside option with a simple expression for the energy share. Nevertheless, this does not prohibit a feasible GMM-type estimator.

Table A.1 briefly summarizes the similarities and differences between Logit and CES. Results hold for their nested counterpart. A primary difference dwells in the latent micro-foundation determining the observed purchase share. The discrete model assumes many independent individuals, making a one-good, one-time purchase choice over several heterogeneous options. A continuous model like the CES assumes that one individual chooses how much to purchase among a set of goods with strictly positive purchases for each.

I choose Nested CES rather than Nested Logit because inside a plant, the purchase decision would not be an aggregation of one-time discrete choice from infinitely many underlying individuals, but instead a discrete-continuous choice made by a few managers responsible for fuel procurement. This is inside the scope of CES. More flexible models would be feasible if the price endogeneity is assumed away, as indicated in Train and Wilson (2011) and Bhat (2005). However, the non-purchased good inside the model makes the Berry (1994)-type inversion invalid, causing insurmountable to account for price endogeneity.

Aspects	**Discrete Choice – Logit**	**Continuous Choice – CES**
Market participants	Infinitely many individuals, each purchase one unit of good	One individual, purchase from multiple sources, multiple units, subject to the budget constraints
Purchase multiple goods	People have different observed or unobserved tastes	Diminishing return to scale, optimal obtained when marginal cost of each good equalized
Estimation Method	Usually parametric, needs to specify the full error distribution	Semi-parametric, extra identification power from the budget constraints
Main Application	Industrial Organization Transportation	International Trade Macro Economics
Similarities	IIA in simple case; same expression of market share; allow for RC and nests; no zero market share permitted.	

Table A.1: Comparison of Logit and CES

Appendix A.2 Nash bargaining related Calculation

For the Nash bargaining problem in (15) and the FOC in (16), Here, by assuming that prices and products' characteristics stay the same without capacity constrained, when bargaining fails, the derivatives $\partial\pi_{mt}(T^*_{-ij}, P^*_{-ij})/\partial p_{ijt} = 0$, $\partial\pi_{it}^{NFC}(T^*_{-ij}, P^*_{-ij})/\partial p_{ijt} = 0$, as in Grennan (2013). Without good j, plant i would re-optimize over all the goods in its current purchase set temporarily regardless of mines' capacities.

And the derivative of profits for coal company m equals

$$\frac{\partial\pi_{mt}(T^*, P^*)}{\partial p_{ijt}} = \left[\sum_{\ell\in J_{imt}} \frac{\partial Q_{i\ell t}(T^*, P^*)}{\partial p_{ijt}}(p_{i\ell t} - mc_{i\ell t})\right] + Q_{ijt}(T^*, P^*)$$

Furthermore, π_{it}^{NFC} is defined as in the main text. The derivative of the NFC surplus for plant i are

$$\frac{\partial\pi_{it}^{NFC}}{\partial p_{ijt}} = \left[AR_{it} - eff_{it} \times AC_{it}\right] \times \frac{\partial Gen_{it}}{\partial AC_{it}} \times \frac{\partial AC_{it}}{\partial p_{ijt}} - Gen_{it} \times eff_{it} \times \frac{\partial AC_{it}}{\partial p_{ijt}} \tag{A.9}$$

The derivative takes a complicated form for both $\partial Gen_{it}/\partial AC_{it}$ and $\partial AC_{it}/\partial p_{ijt}$. AC_{it} is the optimal function value of a constrained optimization, the derivative can be calculated using the Envelope

Theorem with equality constraints, with the Lagrange

$$\mathcal{L} = \Big(\sum_{g_{it} \in G_{it}} \Big[\Big(\sum_{\ell=1}^{J_{g_{it}}} \phi_{i\ell t}^{1-\rho} \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho} \Big)^{\frac{1}{\rho}} \Big]^{\sigma} \Big)^{\frac{1}{\sigma}} + \lambda \Big[1 - \sum_{\ell \in g_{it}, g_{it} \in G_{it}} S_{i\ell t} \Big]$$

And take the derivative as in (Appendix A.1), with the last term equals 0 since the summation of energy share, S_{ijt}, would always be 1 in any feasible solution.

$$\frac{\partial AC_{it}(T^*, P^*)}{\partial p_{ijt}} = \frac{\partial \mathcal{L}}{\partial p_{ijt}} = \frac{\partial}{\partial p_{ijt}} \Big\{ \Big(\sum_{g_{it} \in G_{it}} \Big[\Big(\sum_{\ell=1}^{J_{g_{it}}} \phi_{i\ell t}^{1-\rho} \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho} \Big)^{\frac{1}{\rho}} \Big]^{\sigma} \Big)^{\frac{1}{\sigma}} + \lambda \Big[1 - \sum_{\ell \in g_{it}, g_{it} \in G_{it}} S_{i\ell t} \Big] \Big\}$$

$$= \frac{1}{\sigma} \Big(\sum_{g_{it} \in G_{it}} \Big[\Big(\sum_{\ell=1}^{J_{g_{it}}} \phi_{i\ell t}^{1-\rho} \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho} \Big)^{\frac{1}{\rho}} \Big]^{\sigma} \Big)^{\frac{1}{\sigma} - 1} \times \Big[\sum_{g_{it} \in G_{it}} \Big\{ \sigma \Big[\Big(\sum_{\ell=1}^{J_{g_{it}}} \phi_{i\ell t}^{1-\rho} \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho} \Big)^{\frac{1}{\rho}} \Big]^{\sigma - 1} \times \frac{1}{\rho} \Big(\sum_{\ell=1}^{J_{gi}} \phi_{i\ell t}^{1-\rho} \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho} \Big)^{\frac{1}{\rho} - 1}$$

$$\times \Big[\sum_{\ell=1}^{J_{gi}} \Big(\phi_{i\ell t}^{1-\rho} \times \rho \Big(\frac{S_{i\ell t}}{h_{i\ell t}} \Big)^{\rho - 1} \times \frac{1}{h_{i\ell t}} \frac{\partial S_{i\ell t}}{\partial p_{ijt}} \Big) + (1-\rho) \phi^{-\rho} \times \frac{\partial \phi_{ijt}}{\partial p_{ijt}} \Big(\frac{S_{ijt}}{h_{ijt}} \Big)^{\rho} \Big] \Big\} \Big] - \lambda \sum_{j \in g_{it}, g_{it} \in G_{it}} \frac{\partial S_{i\ell t}}{\partial p_{ijt}}$$

Moreover, there is, $Gen_{it} = \sum_{d \in t} \sum_{o \in d} \sum_{u \in i} G_{udo}$. The derivative of the overall generation of plant i at t regarding the average cost is aggregated from hourly generation,

$$\frac{\partial Gen_{it}}{\partial AC_{it}} = \sum_{d \in t} \sum_{o \in d} \sum_{u \in i} \frac{\partial G_{udo}}{\partial AC_{it}} = \sum_{d \in t} \sum_{o \in d} \sum_{u \in i} \sum_{j=1}^{B} \Big[\mathbf{1}(TG_o \in B_j) \times \frac{1}{p_{NG,d}} \times b_{uj} \Big] \tag{A.10}$$

And the own derivative regarding $Q_{ijt} = E_{it} \times S_{ijt}/h_{ijt}$ equals $\partial Q_{ijt}/\partial p_{ijt} = (1/h_{ijt}) \times [(\partial E_{it}/\partial p_{ijt}) \times S_{ijt} + E_{it} \times (\partial S_{ijt}/\partial p_{ijt})]$.

Appendix B Computational Algorithm

Appendix B.1 Estimation of reservation prices under capacity constraints

For equations like (27) and (29) in the main text, the calculation of the reservation price $p_{ij't}^{RES}$ is required. $p_{ij't}^{RES}$ serves as the minimum acceptable price for the coal company m given the current network T^*, P^*. In other words, this can be regarded as the optimal credible threat extracted from m by i for good j'. With binding capacity constraints, $p_{ij't}^{RES}$ would be larger than the marginal cost and may depend on the rest of the network T^*.

There are two factors making j' a credible threat – profitability and feasibility. Profitability means company m's profit would not fall when its customers switch from i' to i. i' need not be a singleton. For instance, if the opportunity cost of selling to i is the marginal cost, then any price higher than or equal to its marginal cost would be profitable. Moreover, since I summarize a power plant as an optimization machine through two stages of demand, as mentioned in Section 5.2, coal companies cannot specify both quantity and price. Instead, quantity is a function of prices and companies can only influence quantities through prices. Therefore, feasibility refers to the case where the quantities required by plants are still within the capacity constraint. If a $p_{ij't}$ satisfies the profitability and feasibility simultaneously, then $p_{ij't}$ is incentive compatible to coal company m.

However, usually there are more than one incentive compatible prices left for plant i to choose from as a threat. Denote the set of prices as $\mathscr{P}_{ij't}$. Power plants, on the other hand, would pick the $p^{RES}_{ij't} \in \mathscr{P}_{ij't}$ such that

$$p^{RES}_{ij't} = \underset{p_{ij't} \in \mathscr{P}_{ij't}}{\arg\max} \quad \pi_{it}(T^*_{-ij+ij'}, P^*_{-ij+ij'}) - \pi_{it}(T^*, P^*) \tag{B.1}$$

For instance, without capacity constraint, marginal cost is an incentive compatible value at which coal company can sell as much as possible. And it is the global optimal price that a plant can retrieve from the suppliers. Hence, $p^{RES}_{ij't}$ equals to the marginal cost. While situations get more complicated with capacity constraints, algorithms below provide a practicable way to calculate $p^{RES}_{ij't}$.

First, note that, since the prices are determined either by Nash bargaining or the threat of replacement, and Nash bargaining is the upper bound for a feasible price, there is no partial switch of sales from i' to i with a binding capacity constraint. That is, if company m need to curtail the quantities sold to i', then it has to set $Q_{i'j't} = 0$. Remember that quantity is an outcome variable paired with current price level. To reduce the sales in i', $p_{i'j't}$ needs to increase. But an increased price would either cause a break in Nash bargaining, or a replacement. In both cases, the transaction of j' between i' completely dissolve.

Next, I provide algorithms for different plants' surplus. With the variable surplus defined in (11), one can obtain $p^{RES}_{ij't}$ through Algorithm 2-1 below.

Algorithm 2-1. *Regular case – calculate $p^{RES}_{ij't}$ and the optimal surplus of replacement j with j', or as $\pi_{ijj't}$, proceeds on in the following steps:*

i). *For j', get the list for all the current spot sales to $i' \in \mathbb{I}$, given (T^*, P^*). Calculate m's profits for selling j' to i' as $\Gamma_{mi'j't} = \sum_{\ell \in J_{i'mt}} Q_{i'\ell t}(T^*, P^*)(p^F_{i'\ell t} - mc_{i'\ell t}) - \sum_{\ell \in J_{i'mt} \setminus j'} Q_{i'\ell t}(T^*_{-i'j'}, P^*_{-i'j'})(p^F_{i'\ell t} -$*

$mc_{i'\ell t}$) [71]*. Then order the sales records* ***ascendingly*** *by $\Gamma_{mi'j't}$.*

ii). *Set $\tilde{Q}_{ij't} = 0$. $\tilde{Q}_{ij't}$ keeps track of the maximum quantity spared by m for j' that can be purchased by i. Start from the least-profitable sales to $\underline{i}'$. Obtain a $p^F_{ij't} = p^F_{i\underline{i}'j't}$ such that there is*

$$\Gamma_{m\underline{i}'j't} = \sum_{\ell \in J_{imt} \cup j'} Q_{i\ell t}(T^*_{+ij'}, P^*_{+ij'})(p^F_{i\ell t} - mc_{i\ell t}) - \sum_{\ell \in J_{imt}} Q_{i\ell t}(T^*, P^*)(p^F_{i\ell t} - mc_{i\ell t}) \tag{B.2}$$

Equation (B.2) requires that one finds a price $p^F_{i\underline{i}'j't}$ that would gives the same profits as selling to $\underline{i}'$. If one can find such a price, then put $\underline{i}'$ into a set $\mathcal{H}$. Otherwise there is no replacement since the profitability would never be satisfied. Set $\tilde{Q}^1_{ij't} = Q_{\underline{i}'j't}(T^*, P^*)$.

iii). *Suppose there are now k entries in $\mathcal{H}$, with corresponding profit-equivalent prices $p^F_{i\mathcal{H},j't}$. Then proceed on with the next (k+1)th least-profitable sales to i' and obtain $p^F_{ij't} = p^F_{i\{\mathcal{H}\cup i'\}j't}$. If there is no such $p^F_{ij't}$ then go to step* ***iv).***

With a valid profit-equivalent price $p^F_{ij't}$, check whether or not either of conditions in (B.3) are satisfied for $Q_{ij't}$ corresponding to.

$$(1)\ Q_{ij't} \leqslant \tilde{Q}^k_{ij't} + Q_{i'j't}(T^*, P^*),\ \textit{and}\ \ Q_{ij't} \geqslant \tilde{Q}^k_{ij't}; \quad (2)\ Q_{ij't} \geqslant \tilde{Q}^k_{ij't} + Q_{i'j't}(T^*, P^*) \tag{B.3}$$

If satisfied and there is still i'' in the reference list but not in $\mathcal{H}$, set $\mathcal{H} = \mathcal{H} \cup i'$, $\tilde{Q}^{k+1}_{ij't} = \tilde{Q}^k_{ij't} + Q_{i'j't}(T^*, P^*)$ and $k = k + 1$. Then proceed to the next iteration. Otherwise, $\mathcal{H} = \mathcal{H}$ and move forward to step ***iv).***

iv). *Finalize the imputed price $p^{F,*}_{ij't}$. If $Q_{ij't}$ corresponding to $p^F_{ij't}$ is smaller than or equal to $\tilde{Q}_{ij't}$, then $p^{F,*}_{ij't} = p^F_{ij't}$. Otherwise set $p^{F,*}_{ij't}$ as the price which would make $Q_{ij't} = \tilde{Q}_{ij't}$, which must be larger than the profit-equivalent price. With $p^{F,*}_{ij't}$, $\pi_{ijj't}$ follows directly from (11) or (22).*

Algorithm 2-1, in fact, helps to find the minimum credible price that company m offers to plant i for good j'. As adding more and more i' into $\mathcal{H}$, the profit-equivalent price increases and the corresponding $Q_{ij't}$ decreases. When price is low, plant i desires a $Q_{ij't}$ larger than the spared $\tilde{Q}_{ij't}$, which is not feasible. As more quantities add into $\tilde{Q}_{ij't}$, $p_{ij't}$ increases towards the maximum price acceptable to plant i. Algorithm 2-1 stops if and only if a balance is reached. By the surplus defined in (11), a minimum price directly ensures the largest surplus π^{NFC}_{it}.

71. Here, use Γ rather than π to eliminate the possible confusion.

To illustrate why Algorithm 2-1 works, suppose Algorithm 2-1 ends up in a set $\mathscr{H}$, a quantity $\tilde{Q}_{ij't}$ and a reservation price $p_{ij't}^{RES}$. Prices need increase to sell a less than $\tilde{Q}_{ij't}$, which deviates from the optimal surplus of i. Besides, Algorithm 2-1 stops when (1) cannot find a profit-equivalence price in $\mathscr{H} \cup i'$; (2) as implied by condition B.3, $Q_{ij't} \leqslant \tilde{Q}_{ij't}^{k}$; or (3) there does not exist a $i' \notin \mathscr{H}$. For (1), having a lower price and a larger quantity is not possible if more i' adds into $\mathscr{H}$; therefore, the current solution is the minimum. For (3), which almost never happens in real data, there is no more capacity to add in, so the minimum price is the price corresponding to selling the entire capacity of j' to i. In case (2), no more capacity is needed.

If I have, as it is in this study, (22) rather than (11), the monotonic relation between $p_{ij't}$ and $\pi_{ij't}$ collapses. Further steps in Algorithm 2-2 deal with this.

Algorithm 2-2. *Calculate $p_{ij't}^{RES}$ and the optimal surplus of replacement j with j', or $\pi_{ijj't}$, for surplus defined in (22). Proceed on in the following steps:*

i). *Start by the* ***ascendingly****-sorted sales list as in Algorithm 2-1.*

ii). *Calculate the $(\hat{p}_{ij't}, \hat{Q}_{ij't})$ pair from Algorithm 2-1.*

iii). *Then check whether or not $(\hat{p}_{ij't}, \hat{Q}_{ij't})$ leads to a postive $\sum_{J'_{it}}\{\max\{0, \underline{Q}_{ijt} - Q_{ijt}\} \times p_{ijt}^{F}\}$. If not, $p_{ij't}^{RES} = \hat{p}_{ij't}$. Otherwise, $\hat{p}_{ijt}$ may be too low, if $\partial\pi_{it}^{NFC}/\partial p_{ij't}$ is positive when evaluated at $\hat{p}_{ij't}$. With this, increase $p_{ij't}$ until $\partial\pi_{it}^{NFC}/\partial p_{ij't} = 0$.*

The adjustment in Algorithm 2-2 iii) for is both profitable and feasible, since the price becomes higher than a profitable and feasible solution in Algorithm 2-1. Algorithm 2-2 iii) accommodating (22) by giving a local modification towards the maximum of plant i's surplus under the non-monotonicity of π_{it}^{NFC} regarding $p_{ij't}$.

Appendix B.2 Proofs regarding Algorithm 1

First, let me show the justification of transforming equality constraint (25) to the partitioned-out $f_{ijt} = \min\{f_{ijt}^{Nash}, \min_{j' \text{ with } Q_{ij't}=0}\{f_{ij't} - f_{ijj't}^{diff}\}\}$, $\forall j$ with $Q_{ijt} > 0$. For a purchased good j, if f_{ijt}^{Nash} satisfies the no-replacement condition, or $f_{ijt}^{Nash} \leqslant \tilde{f}_{ijj't} = f_{ij't} - f_{ijj't}^{diff}$. Suppose in this case, $f_{ijt} = \tilde{f}_{ijj't} \geqslant f_{ijt}^{Nash}$, then the observed p_{ijt} cannot be generated by the Nash bargaining. However, according to the conditions for a stable network, a rational seller would set the price equals to the Nash bargaining price unless it is not compatiable for the replacement. In other words, the prices implied from no-replacement would be lower than Nash bargaining price given the same f_{ijt}.

To check this, consider the equation (17) in the main text. There is an increase from f_{ijt}^{Nash} to $\tilde{f}_{ijj't}$. To keep a equality, the right side needs to increase too. Prices need to decrease since in a regular case, there are . Thus, under $\tilde{f}_{ijj't}$, $p_{ijt}^{Nash} \leqslant p_{ijt}$, which sets out the contradiction.

Next, I show some extra proofs for the alternating method for the inner loop in Algorithm 1. First, the inner loop (30) is a downhill descending method that the objective function value always decreases from iteration k to k + 1. To see this, suppose currently I have $\{f_{ijt}^k, \chi^k, FE_j^k, FE_t^k\}$, $\forall i, j, t$. Given f_{ijt}^k, solve the linear regression such that,

$$\{f_{ijt}^{k+1}, \chi^{k+1}, FE_j^{k+1}, FE_t^{k+1}\} = \arg\min g_1^{k+1} = \arg\min \sum_{i,j,t} [f_{ijt}^k - FE_j - FE_t - \chi\Omega_{ijt}]^2 \quad \text{(B.4)}$$

The function value g_1^{k+1} in (B.4) must be $\leqslant g^k$, because $\{\chi^k, FE_j^k, FE_t^k\}$ is a feasible solution to (B.4). Similar logic applies to the f_{ijt}-updating step. Since the constraints $\underline{f}_{ijt}$, f_{ijt}^{Nash} and $f_{ijj't}^{diff}$ are determined outside the system, and whether or not f_{ijt}^k satisfies the constraints does not depend on the linear parameters, f_{ijt}^k is also feasible for problem (30), so that the value function g_2^{k+1} matched with f_{ijt}^{k+1} would have $g_2^{k+1} \leqslant g_1^{k+1}$. Moreover, objective function g is the summation of squared terms, which is bounded below by zero. Therefore, the inner loop in Algorithm 1 would converge to a fixed point, denoted as $\{\check{f}_{ijt}, \check{\chi}, \check{FE}_j, \check{FE}_t\}$.

The fixed point $\{\check{f}_{ijt}, \check{\chi}, \check{FE}_j, \check{FE}_t\}$ is a local minima for problem in $(*)$. Global minima may be attainable with different starting values. In contract, Ghili (2022) sacrifice the exact equality and replace (25) in $(*)$ with $f_{ijt} \leqslant f_{ijt}^{Nash}$, $f_{ijt} \leqslant f_{ij't} - f_{ijj't}$, $\forall Q_{ij't} = 0$, $\forall Q_{ijt} > 0$. In fact, Ghili (2022) attains global convergence and a fast computational speed in exchange for a compatibility of no-replacement, while I resort to the opposite in this study.

Appendix B.3 Algorithm for Counterfactual

The algorithm I embraced for the counterfactual extends from Fan and Yang (2020). I start from a random network T^0, generate the corresponding prices P^0, and then only check the networks reached from a sequence of unilateral deviations. For simplicity, T^0 is the observed transactions, and P^0 the corresponding prices.

The counterfactual algorithm I adopt here can be found in Algorithm 3, and a graphic illustration in Figure C.3. Denote the set of all the goods at t as $\mathbb{J}^0_{it}$. $\mathbb{J}^0_{it}$ contains all inside goods in PRB, the composite other PRB coal, and all other non-PRB coal.

Algorithm 3. *One can calculate a new stable network (T^*, P^*) locally from a starting point (T^0, P^0) by the following steps:*

i). *Initialization – start from a network (T^0, P^0).*

ii). ***Outer loop*** *starts. Set $\mathbb{J}_{it} = \mathbb{J}^0_{it}$.*

iii). *For the network of current iteration (T^d, P^d), pick a coal mine $j \in \mathbb{J}_{it}$ from a per-determined order. Then obtain the set of plants $\mathbb{I}^d$ whose profits would increase with addition or cut-off.*

iv). *Choose $i \in \mathbb{I}^d$ with the largest Ξ_i after a one-step calculation of the best response prices. Ξ_i is defined as a weighted gain of the direct participants after the network change,*

$$\Xi_i = \omega_\Xi \times [\pi_{it}(T^d, P^d) - \pi_{it}(T^d_{\pm ij}, P^d_{\pm ij})] + (1 - \omega_\Xi) \times [\pi_{mt}(T^d, P^d) - \pi_{mt}(T^d_{\pm ij}, P^d_{\pm ij})] \qquad \text{(B.5)}$$

v). ***Inner loop*** *starts. Solve the best response price for the chosen i. Then check whether the new change in the network is incentive compatible under current prices by examining,*

$$\begin{aligned}
&\textit{For addition:} \quad \pi_{it}(T^d, P^d) - \pi_{it}(T^d_{+ij}, P^d_{+ij}) \geqslant 0,\ \pi_{mt}(T^d, P^d) - \pi_{it}(T^d_{+ij}, P^d_{+ij}) \geqslant 0,\\
&\textit{For single cut-off:} \quad \pi_{mt}(T^d, P^d) - \pi_{it}(T^d_{-ij}, P^d_{-ij}) \geqslant 0,\\
&\textit{For replacing cut-off:} \quad \pi_{mt}(T^d, P^d) - \pi_{it}(T^d_{-ij}, P^d_{-ij}) \geqslant 0,\\
&\qquad\qquad\qquad\qquad\quad \pi_{it}(T^d, P^d) - \pi_{it}(T^d_{+i'j}, P^d_{+i'j}),\ \forall i' \textit{ with } Q_{i'jt} > 0.
\end{aligned} \qquad \text{(B.6)}$$

*If (B.6) is satisfied, then set $T^{d+1} = T^d_{\pm ij}$, $d = d+1$, $\mathbb{J}_{it} = \mathbb{J}_{it} \backslash j$. Otherwise set $\mathbb{I}^d = \mathbb{I}^d \backslash i$. If $\mathbb{I}^d \neq \varnothing$, go back to step **iv)**. Otherwise go back to step **iii)**.*

vi). *Check if $\mathbb{J}_{it} = \varnothing$. If not, start a new iteration of step **ii)**. Otherwise terminate the algorithm with a stable network $(T^* = T^d, P^* = P^d)$.*

Algorithm 3 has an intuitive explanation. A stable network is a synergy where demand and supply sides join together, with the inner loop tackling the demand side and the outer loop for the supply side. In the inner loop, prices are set so that no power plant wants to deviate from the current network. Those best response prices exist for a random network structure but do not account for the seller's actions. Outer loop searches over networks to satisfy coal companies' incentive compatibility given the best response price.

The rest of this section discusses some numerical challenges during the counterfactual. Firstly, the counterfactual results depend on the fixed order of $j \in \mathbb{J}_{it}$ in **ii)**, which is chosen without any theoretical guidance, especially when there is a tight capacity constraint. To overcome this randomness, I use three different orders [72] and get set approximation of counterfactual prices. Secondly, Algorithm 3 may not converge to any stable network. Infinite loop may occur before reaching a stable network by having a few links added and deleted constantly. For this, either I would change the order of $j \in \mathbb{J}_{it}$ temporarily to recover, or I may ignore this attempt. Thirdly, another tuning parameter in Algorithm 3 is the weight ω_{Ξ} in (B.5). Step **iv)** gives a preliminary screening [73]to choose the plant with the largest increase increase in the weighted social welfare. For instance, $\omega_{\Xi} = 0.5$ means an incremental in consumer's surplus and seller's profits are treated equally. For robustness, I vary the ω_{Ξ} and repeatedly run the Algorithm 3.

Even for a given network structure T^d at hand, numerically calculating prices and quantities of the inside goods satisfying both the mutual best response is not an easy task. Here I re-state the fixed-point iteration I use to solve P^d. Recall in the main text, the price p_{ijt} is defined as $p_{ijt} = \min\{p_{ijt}^{Nash}, p_{ijt}^{NR}\}$. Both p_{ijt}^{Nash} and p_{ijt}^{NR} are functions of T^d and the rest of prices $\boldsymbol{P}_{-ijt}$[74]. Holding $\boldsymbol{P}_{-ijt}$ constant, I can calculate p_{ijt}. Prices would reach a mutual best response when it comes to a fixed point.

Suppose there is a vectorized function ϕ. For each j, it calculates the price holding the rest constant. Then a mutually best response price P^d is the solution for the system of equations that,

$$\boldsymbol{P}_t = \phi(\boldsymbol{P}_t) \tag{B.7}$$

72. Orders (1): the good with largest absolute violation of capacity constraints moving first; (2): the good with largest proportional violation of capacity constraints moving first; (3): a fixed sequence based on the overall traded volumes, and keep still across Algorithm 3.

73. It is preliminary since it comes from one-step best response, which is a compromise. Calculating fixed point or k-step best response prices for all $i \in \mathbb{I}_{it}$ is time-consuming.

74. For a valid Nash bargaining price, I need to have $\pi_{it}^{NFC}(T^*, P^*) - \pi_{it}^{NFC}(T^*_{-ij}, P^*_{-ij}) - f_{ijt} \geqslant 0$. Nevertheless, in numerical calculation, for a network-price combination, sometimes the unconstrained p_{ijt}^{Nash} may not be calculated correctly. It may deviate to an unreasonably large value. To make the algorithm robust, in practice, I include an extra term $p_{ijt}^{nocut-off}$ in practice,

$$p_{ijt} = \min\{p_{ijt}^{Nash}, p_{ijt}^{NR}, p_{ijt}^{no-cutoff}\}$$

$p_{ijt}^{no-cutoff}$ is the price that makes $\pi_{it}^{NFC}(T^*, P^*) - \pi_{it}^{NFC}(T^*_{-ij}, P^*_{-ij}) - f_{ijt} = 0$. $p_{ijt}^{no-cutoff}$ would act as a soften cushion against possible failure of p_{ijt}^{Nash}, allowing the fixed-point iteration in (B.7) to proceed on. This would help if the failure is temporary due to a poor starting or intermediate value.

And the parallelable Gauss-Jacobi iteration is employed for the fixed-point iteration in B.7 such that,

Algorithm – Gauss-Jacobi. *To find the fixed-point solution to (B.7).*

i). *Given the current iteration of* $\boldsymbol{P}_t^k$*, at k-th iteration;*

ii). *Parallel computing: for each i, j, calculate the new iteration* $p_{ijt}^{k+1} = \min\{p_{ijt}^{Nash}, p_{ijt}^{NR}\}$*, where both* p_{ijt}^{Nash} *and* p_{ijt}^{NR} *is obtained by holding* $\boldsymbol{P}_{-ijt}^k$ *fixed.*

iii). *Update* $\boldsymbol{P}_t^{k+1}$ *using the* p_{ijt}^{k+1} *calculated in step* ***ii).*** *Set* $k = k + 1$*.*

iv). *Continue from* ***i)*** *to* ***iii)*** *until convergence is reached, or some stopping criterion is met. Convergence is measured by* $||\boldsymbol{P}_t^{k+1} - \boldsymbol{P}_t^k||$*.*

An example of the non-convergence stopping criterion in algorithm above is the maximum iterations K. This K-step method is justified in, for example, Lee and Seo (2015) for Berry et al. (1995)-type models. Terminating after K steps would reduce the computational time. For some poor starting values, it prevents the $\boldsymbol{P}_t$ from slipping into unrealistic values. Therefore, I start the Algorithm 3 with a large K to keep a precision estimation. Then decrease K in the middle to reduce the elapsed time. Finally gradually increase $K \to \infty$ to get the fixed solution.

Another computational issue arises when the number of solid links increases to roughly larger than thirty. Gauss-Jacobi iteration then becomes hard to converge. Instead of fixed-point convergence, it may oscillate back and forth between two price levels [75]. To alleviate this problem, I take the step implied by the spectral algorithm in Barzilai and Borwein (1988) and La Cruz et al. (2006), whose effectiveness in dealing with unstable fixed points in dynamic discrete choice NPL method has been demonstrated in Aguirregabiria and Marcoux (2021). The updated $\boldsymbol{P}_t^{k+1}$ is becomes a dampened step following a linear combination of the calculated new iteration $\phi(\boldsymbol{P}_t)$ and the current residuals, $\boldsymbol{P}_t^{k+1} = \boldsymbol{P}_t^k - \alpha_k(\phi(\boldsymbol{P}_t^k) - \boldsymbol{P}_t^k)$.

The scalar α_k is calculated following Barzilai and Borwein (1988),

$$\alpha_k = \frac{(\boldsymbol{P}_t^k - \boldsymbol{P}_t^{k-1})'[\phi(\boldsymbol{P}_t^k) - P_t^k - (\phi(\boldsymbol{P}_t^{k-1}) - \boldsymbol{P}_t^{k-1})]}{[\phi(\boldsymbol{P}_t^k) - P_t^k - (\phi(\boldsymbol{P}_t^{k-1}) - \boldsymbol{P}_t^{k-1})]'[\phi(\boldsymbol{P}_t^k) - P_t^k - (\phi(\boldsymbol{P}_t^{k-1}) - \boldsymbol{P}_t^{k-1})]} \tag{B.8}$$

75. It is a purely numerical problem, with some fixed points less numerically stable than others. It does not refer to the non-existence of a fixed point.

and La Cruz et al. (2006),

$$\alpha_k = \frac{(\boldsymbol{P}_t^k - \boldsymbol{P}_t^{k-1})'(\boldsymbol{P}_t^k - \boldsymbol{P}_t^{k-1})}{(\boldsymbol{P}_t^k - \boldsymbol{P}_t^{k-1})'[\phi(\boldsymbol{P}_t^k) - P_t^k - (\phi(\boldsymbol{P}_t^{k-1}) - \boldsymbol{P}_t^{k-1})]} \tag{B.9}$$

One can see that whenever $\alpha_k = -1$, a full step is taken by having $\boldsymbol{P}_t^{k+1} = \boldsymbol{P}_t^k$. But if $\alpha_k \neq -1$, then it incurs an adjustment by approximating the derivatives using a derivative-free, simply-calculated analytical value. This type of derivative-free methods is a competitive algorithm in solving moderate-to-high scale nonlinear equation problem. The algorithm terminates whenever the original maximum absolute distance $\boldsymbol{P}_t - \phi(\boldsymbol{P}_t)$ falls below some threshold.

Appendix C Miscellaneous on Model Setting and Estimation

Appendix C.1 More on diff-in-diff

Here are some complementary results of the matched diff-in-diff regressions. Table C.1 to Table C.2 show the results when the problematic periods are removed from the sample. γ_1 is absorbed into the Year fixed effect, so is omitted whenever Year fixed effect presents. In general, the diff-in-diff setting in the main text is robust to the set of data I use.

Appendix C.2 Level-price or log-price in CES

In the structural demand estimation, I choose to use level price rather than the log price. In contrast with the widely-used log price, in this study, level price supports a more reasonable counterfactual behavior. To illustrate, for a randomly-chosen power plant, plant # 177, and good # 1, and also for a bunch of candidates of price coefficient, I increase and decrease the price level until it goes beyond the range of prices observed in the transaction data. Figure C.1 displays the corresponding optimal purchased shares. For the log-price case (panel (a)), with price coefficient equals to -1.15, the plant # 177 would still have around 10% of energy share relies on good # 1 when price is as high as 200. It is hard to think about a situation where an average plant wants to purchase a fair amount of coal at five times of the prevailing price.

In the opposite, the market share decays fast with the level price when prices get abnormally high. Even though plant # 177 finds good # 1 very attractive through some unobservables, it is still not tolerant towards extraordinary expenses. Moreover, shares also increase very fast with a

	Matched		Not Matched	
Deregulation	-0.3023***	-0.2517**	-0.1034***	-0.0164
	(0.1009)	(0.1022)	(0.0386)	(0.0392)
γ_1	-	0.1816***	-	0.1788***
	(.)	(0.0594)	(.)	(0.0149)
γ_2	-0.0342	-0.0366	-0.0310	-0.0359
	(0.0338)	(0.0340)	(0.0224)	(0.0230)
γ_3	0.1572***	0.1560***	0.1417***	0.1389***
	(0.0416)	(0.0417)	(0.0408)	(0.0416)
γ_4	-0.0131	-0.0032	0.2231***	0.2315***
	(0.0915)	(0.0906)	(0.0671)	(0.0685)
Year-FE	Yes	No	Yes	No
Month-FE	Yes	No	Yes	No
vce	cluster	cluster	cluster	cluster
r2	0.9018	0.8994	0.8178	0.8063

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

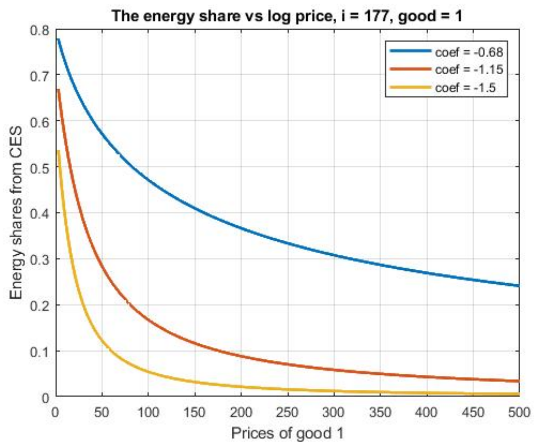
Table C.1: Matched Diff-in-diff, in log price, with two periods removed.

	Quantity Weighted		Not Weighted	
Deregulation	-0.1651***	-0.1294**	-0.0514*	0.0360
	(0.0547)	(0.0516)	(0.0287)	(0.0287)
γ_1	-	0.2767***	-	0.2265***
	(.)	(0.0848)	(.)	(0.0186)
γ_2	-0.1184***	-0.1211***	-0.0880***	-0.0935***
	(0.0381)	(0.0387)	(0.0117)	(0.0122)
γ_3	0.0772***	0.0770***	0.0647***	0.0628***
	(0.0198)	(0.0199)	(0.0170)	(0.0177)
γ_4	-0.0733	-0.0680	0.1062**	0.1140**
	(0.0792)	(0.0795)	(0.0430)	(0.0442)
Year-FE	Yes	No	Yes	No
Month-FE	Yes	No	Yes	No
vce	cluster	cluster	cluster	cluster
r2	0.9046	0.9023	0.9379	0.9328

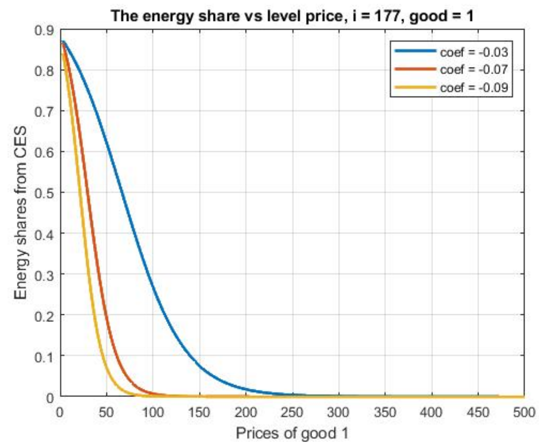
Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table C.2: Matched Diff-in-Diff, in real level price, with two periods removed.

rapidly decreasing price. Hence, level price is more realistic and generate patterns compatible to the observed behaviors for power plants, which is important especially for the counterfactual.



(a)



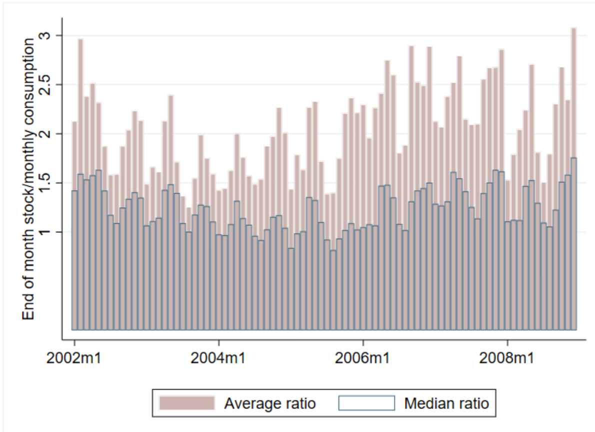
(b)

Figure C.1: Change in purchased share, with log-price or level-price in the structural demand

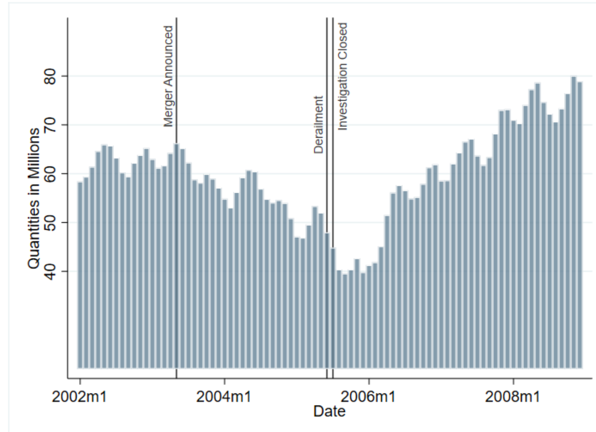
Appendix C.3 Power plants' stockpiling

Power plants store coals on site, which is a potential hazards of my static setting. To check for it, data in EIA-906/920 records quantities, stocked at the end of each month from 2002 to 2008. The data is type-specific, separately for bituminous, subbituminous, and lignite. For simplification, I aggregate different types together for Figure C.2, panel (a). Roughly an average plant stores the coals for 1.5-to-3-month consumption. A median plant invests even less in storage, indicating that the average value may be push up by some extreme plants that store a huge number of coals. There are large variations among plants. The most precautious plants may keep enough coals for six months' consumption, while mine-mouth plants may find no necessities to store at all.

Rigorously, this stockpiling behavior needs to be accounted for. Unfortunately, a dynamic model explicitly describing the stockpiling choice, like in Jha (2019), is computationally too difficult. Therefore, I just adopt a heuristic way and add the monthly transaction data to the quarter-level. Three-month time period is wide enough for most plants to finish the entire purchase-consumption procedure. Further aggregation into half a year or one year is problematic, since choice sets may vary through a year. For example, some coal sources may be available in one or two months at an excellent price, but that price may not last for long. Consequently, with yearly aggregation, the price-quantity relationship is probably reversed and plants may favor higher-priced goods instead.



(a) The average and median value of quantity stocked divided by monthly consumption.



(b) The stock of PRB coal

Figure C.2: On power plants' stockpiling.

Appendix C.4 The full set of parameters estimated in Stage II

Table C.3 summarizes the parameters in the stage II demand estimation. Besides the price, and the diminishing return to scale parameters, I also include the product characteristics accessible in Platts data. Those characteristics include the sulfur content adjusted by the price for sulfur allowance, ash content, and the proportion of generating units with sulfur-capturing scrubbers installed. Also I include the interactions of long-term contract indicator with the those characteristics, and a rich set of dummy variables, which is even more crucial in the pairwise different setting. Pairwise assumes that the errors ξ_{ijt} and $\xi_{ij't}$, though from different goods, their conditional mean is the same. The column "Estimated results" in Table C.3 corresponds to Model 4 in Table 6.

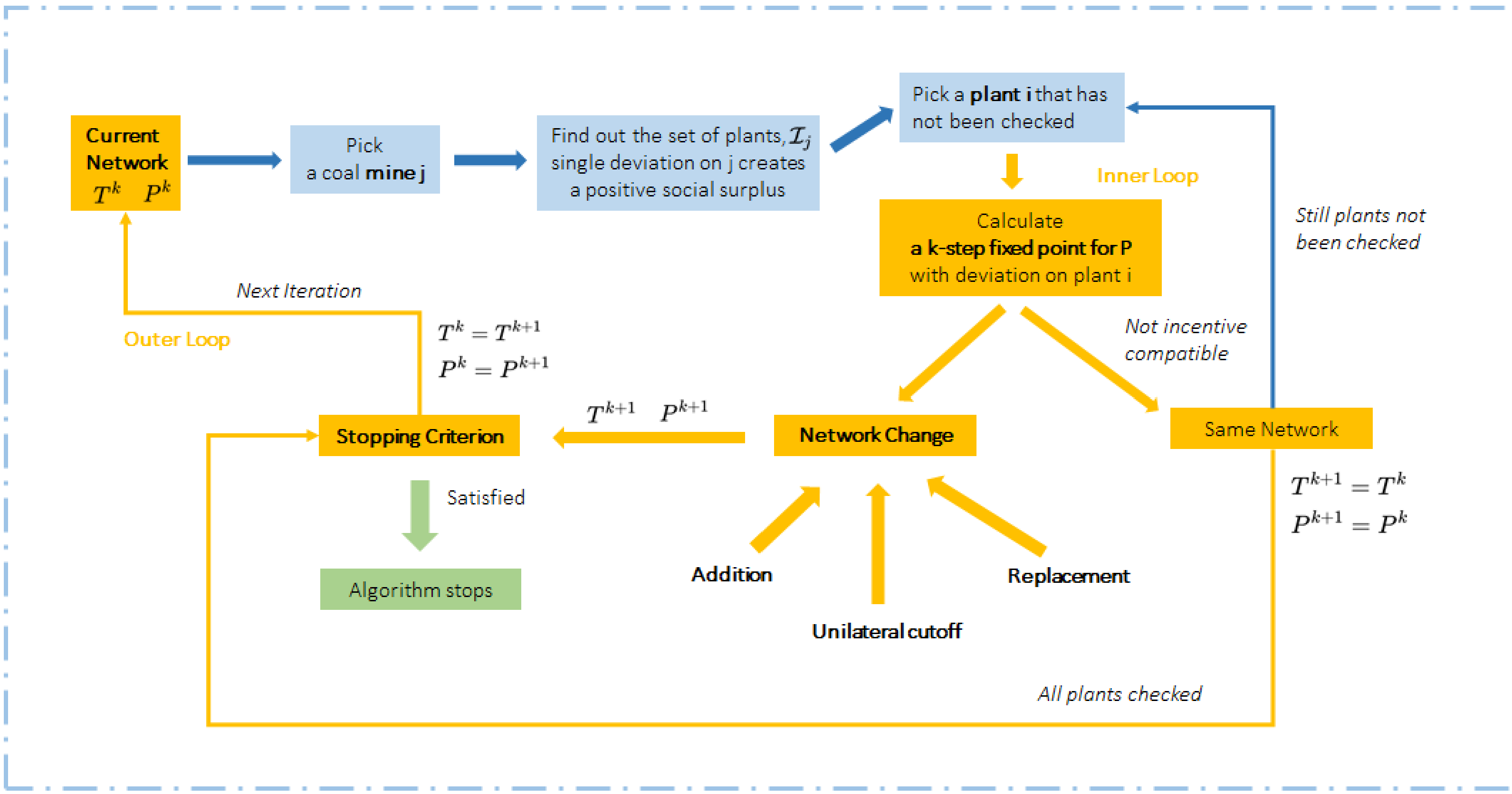


Figure C.3: Flowchart for Algorithm 3.

Name	Description	Expected sign	Estimated results
Sulfur	Sulfur content × Price of sulfur emission allowance	–	–
Ash	% of Ash content	–	–
Price	Real price. $p_{ijt} = p^F_{ijt} = p^T_{ijt}$	–	–
Sulfur × Scrub	Sulfur × (# of generating units/# of scrubbers)	+	+
Deregulation × Price	$\mathbf{1}\{$i is a deregulated plant$\}$	+	
Distance	Haversine distance calculated from latitude & longitude.	–	+
σ	Diminishing return to scale.	> 1	> 1
ρ		$\rho > 1$, $\rho < \sigma$	$\rho > 1$, $\rho < \sigma$
Price × LR	Characteristics interacting with long-term contract indicator	+	+
Sulfur × LR		Any	–
Ash × LR		Any	+
Sulfur × Scrub × LR		Any	+
Distance × LR		Any	–
Dummies	Major Supply Region × Year-Quarter × contract type.	Any	

Table C.3: The Parameters included in the Demand-Stage II Nested CES estimation – a brief description.