American Economic Review 2021, 111(4): 1356–1372
https://doi.org/10.1257/aer.20200679

When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation: Comment†

By Jin Soo Han, Jean-François Houde, Arthur A. van Benthem, and Jose Miguel Abito*

We revisit one of the results in Cicala (2015) and show that the previously estimated large and significant effects of US electricity restructuring on fuel procurement are not robust to the presence of outliers. Using methodologies from the robust statistics literature, we estimate the effect to be less than one-half of the previous estimate and not statistically different from zero. The robust methodology also identifies as outliers the plants owned by a single company whose coal contracts were renegotiated before discussions about restructuring even started. (*JEL* D83, O13, O33, Q16)

The regulatory restructuring of the US electricity generation market, initiated in the late 1990s, represents one of the most ambitious efforts by a government to provide market-based incentives to reduce costs (Laffont and Tirole 1993). In many states, regulated utilities were forced to divest their power generating assets, which would then compete in a competitive wholesale market. The impact of these reforms has been studied extensively, and the literature has identified several channels through which divestiture has improved firms' overall performance: productivity gains (Fabrizio, Rose, and Wolfram 2007), more efficient investments (Fowlie 2010), and lower fuel procurement costs (Cicala 2015; Chan et al. 2017; Jha 2019).[1]

Fuel procurement costs are the largest part of power plants' operating expenses, and utility operators play an important role in the procurement process. Given the large role of fuel procurement in driving electricity generation costs, documenting credible and precise estimates of the effect of restructuring on fuel costs is of first-order importance not only to scholars interested in electricity market design,

* Han: Republic of Korea Air Force Academy (email: han3751@gmail.com); Houde: Department of Economics, University of Wisconsin-Madison, and NBER (email: houde@wisc.edu); van Benthem: The Wharton School, University of Pennsylvania, and NBER (email: arthurv@wharton.upenn.edu); Abito: The Wharton School, University of Pennsylvania (email: abito@wharton.upenn.edu). Liran Einav was the coeditor for this article. We thank Steve Cicala, Uli Doraszelski, Meredith Fowlie, Joe Harrington, Akshaya Jha, Doug Miller, Mikkel Sølvsten, and Katja Seim, as well as seminar participants at the University of Pennsylvania, for helpful comments and suggestions. We obtained disclosure approval from the Energy Information Administration to publish this paper. Van Benthem thanks the Wharton Dean's Research Fund for support.

† Go to https://doi.org/10.1257/aer.20200679 to visit the article page for additional materials and author disclosure statements.

[1] Similar efficiency gains have been found in other countries and industries, for instance, Newbery and Pollitt (1997), Wolfram (1999), Malik et al. (2011), and Davis and Wolfram (2012).

but also to a wider audience interested in the effects of deregulation and incentives. Cicala (2015) is a key contribution in the literature as it provides the first estimate of the average "treatment" effect of restructuring on coal procurement costs. The paper uses the quasi-experimental variation induced by the fact that a large number of states stopped their restructuring efforts in 2000 after the electricity crisis in California. This is a particularly credible identification strategy since all states considered the divestiture option by 1998, which implies that non-divested plants can serve as a plausible control group (at least conditional on observable characteristics). The study carefully checks the results from two standard tests of difference-in-differences (DiD) models, parallel pre-trends and balance on observables, none of which raise any obvious concerns.[2] One of the key findings is that coal-fired power plants, on average, achieved a sizable cost reduction of about 12 percent upon divestiture.

In this paper, we apply robust regression techniques to the same DiD research design as in Cicala (2015) and find that outlier events are disproportionally impacting this estimate.[3] We show that the average treatment effect of divestiture on the procurement cost is between -3 percent and -5 percent, and is not statistically different from zero. To obtain this result, we implement a two-step methodology designed to identify influential observations. In the first step, we estimate the treatment effect parameter using an estimator robust to the presence of outliers. In the second step, we use the estimated residuals from the robust regression and apply a threshold rule to flag individual plants who experienced abnormal residual changes before versus after the treatment (Rousseeuw and Leroy 1987).

Invalid outliers, in general, can be caused by data errors or misspecification of the model. In either case, the presence of outlier residuals suggests that individuals are not drawn from a common distribution in the treated and control groups, which makes it difficult to construct a credible counterfactual simply based on matching on observable covariates. The burden of justifying leaving in such outlier observations should be on the researcher, who is well advised to put more emphasis on robust estimators.

In our context, we find that the gap between the robust and least squares estimate is explained by the presence of a group of seven outlier plants in the treatment group belonging to a single utility in Illinois, Commonwealth Edison (henceforth, ComEd,). This finding alone is important for assessing the external validity of the estimate.

Over the full sample, ComEd plants reduced their delivery price of coal by a very large amount, roughly two-thirds of which took place before its divestiture. This price decline corresponds to about four times the standard deviation of (residualized) prices. After studying the (very unusual) contracting history of ComEd, we conclude that this outlier change was in large part caused by a contract renegotiation event that took place in 1992, years before the first hearing about deregulation in Illinois. We show that this change in procurement had long-term effects on the

[2] Online Appendix Section A.1 shows the results from a balance of covariates test and a pre-trends test.

[3] The focus of our analysis is on the average treatment effect of divestiture on coal procurement costs; we do not comment on the other results of the original paper.

delivery price and quantity, and led to the early termination of one of ComEd's two main contracts.

We argue that this outlier should not be used to identify the causal effect of divestiture. First, ComEd's fuel cost was several times larger than for any other plants at the beginning of the sample. This makes the decline between 1992 and 1999 stand out relative to observations in the control group, and implies that the least squares estimate is disproportionally impacted by the presence of ComEd in the sample. Second, the 1992 renegotiation caused a pre-policy trend for the ComEd plants that is unrelated to its eventual divestiture. Third, we provide a series of institutional details regarding the 1992 renegotiation that suggest that ComEd's fuel costs would likely have continued to decline beyond 2000. All three reasons imply that constructing a counterfactual price path for ComEd absent of deregulation is quite difficult, because no control group plant experienced a similar contracting history. In the last section, we present alternative approaches to account for the ComEd-specific trend when estimating the average treatment effect.

The paper makes two contributions. First, we contribute to the economic regulation literature by providing a credible and robust estimate of the effect of divestiture on procurement cost savings, which challenges the conventional wisdom that divestiture led to large fuel cost reductions. Our results show that deregulation did not provide substantial additional incentives to the average utility in negotiating contracts with mines.[4] This is consistent with the possibility that prior incentive regulation and monitoring by regulatory agencies had been effective at reducing fuel cost for the average plant.

Our second contribution is to the broader applied econometrics literature. Robust regression methods are under-utilized in economics, and we demonstrate their usefulness in the context of the estimation of treatment effects. There are growing concerns in economics about the external validity and reproducibility of treatment effect estimators. Recent papers have investigated the impact of outliers on the credibility of estimates. For instance, using Monte Carlo simulations, Ontiveros, Canavire-Bacarreza, and Castro (2017) show how different types of outliers can bias the estimates from matching estimators of treatment effects. Young (2020) performs a comprehensive study of published work and shows that, in a substantial fraction of papers, statistical significance of the IV estimates is largely driven by one or two outlier observations.[5] We see the use of robust regression methods in the context of quasi-experiments as useful tools that complement our standard methods, are easy to use, and can improve the credibility of empirical research.

[4] In ongoing work, we show that *some* plants did in fact negotiate better terms because of divestiture, especially if they were stuck with an unfavorable long-term contract, were not exposed to incentive regulation prior to divestiture, and/or had more bargaining leverage relative to mines (Han et al. 2020).

[5] Outliers also render the standard linear instrumental variables model inefficient or even inconsistent. See Sølvsten (2020) for a related application of robust regression in the linear IV model.

I. Robust Estimation of Treatment Effects

A. *Research Design*

Our main estimating equation is a fixed-effect regression with a treatment indicator variable identifying divested plants:

$$y_{jt} = \log(price)_{jt} = \alpha \mathbf{1}(divest)_j \cdot \mathbf{1}(t \geq t_j^*) + \gamma_j + \delta_t + \eta_{jt}, \tag{1}$$

where $\log(price)_{jt}$ is the logarithm of the coal price (i.e., the price of the coal delivery) for plant j in year-month t in dollars per MMBtu (nominal price divided by heat content), $\mathbf{1}(divest)_j$ is an indicator variable for divested plants, and $\mathbf{1}(divest)_j \cdot \mathbf{1}(t \geq t_j^*)$ is the treatment variable equal to 1 after the divestiture date t_j^*. The parameter α then measures the treatment effect of interest.

Borrowing the notation from Conley and Taber (2011), the OLS estimate of α can be written as (assuming for simplicity the panel is balanced):

$$\hat{\alpha}_{did} = \alpha + \frac{1}{J_1}\sum_{j=1}^{J_1} w_j - \frac{1}{J_2}\sum_{j=J_1+1}^{J_1+J_2} w_j = \Delta\bar{y}_1 - \Delta\bar{y}_2, \tag{2}$$

where J_1 and J_2 refer to the number of plants in treatment and control groups respectively, $w_j = \left(1/(T - t_j^* + 1)\right)\sum_{t=t_j^*}^{T}\eta_{jt} - \left(1/(t_j^* - 1)\right)\sum_{t=1}^{t_j^*-1}\eta_{jt}$ denotes the *long-run* change in the average residual for plant j between the pre- and post-treatment periods, and $\Delta\bar{y}_1 = (1/J_1)\sum_{j=1}^{J_1}\left[\left(1/(T - t_j^* + 1)\right)\sum_{t=t_j^*}^{T} y_{jt} - \left(1/(t_j^* - 1)\right)\sum_{t=1}^{t_j^*-1} y_{jt}\right]$ is the average long-term difference in outcomes among individuals in the treatment group (the control group average difference is defined analogously). We refer to this long-run difference as the *long-difference* throughout the paper.

Outliers can affect both the magnitude of the treatment effect, as well as its standard error. It is well known that least squares estimates such as $\hat{\alpha}_{did}$ in equation (2) are sensitive to outliers because a single observation can bias their values by an arbitrary amount.[6] These observations are labeled as "influential," since they have a disproportionate influence on the estimated regression line and therefore on the treatment effect. In a DiD framework where the treatment variable is discrete, outliers are caused by a *vertical* shift in the outcome variable. From equation (2), we can see that the OLS estimate of α is sensitive to the presence of an individual, or a group of individuals, experiencing abnormal and persistent price changes over the entire treatment period; or equivalently abnormal realization of the aggregate residual w_j. Due to the panel nature of the data, such outliers can be caused by one or a few isolated monthly observations, or, more likely, by a sequence of correlated residuals η_{jt}. This leads to an inference problem. With correlated residuals, outliers occur at the cross-sectional unit level j (or a cluster of such units), rather than at the level of a single observation at the unit-time level jt. This should guide the choice

[6] Formally, sample averages are said to have a *breakdown value* of $1/n$ since a single observation can move the estimate to plus or minus infinity. See Hubert, Rousseeuw, and Van Aelst (2008) for a review of the robust statistics literature.

of clustering for standard errors, which will impact standard error estimates in the presence of outliers.[7]

The statistics literature on robust estimation and outlier detection recommends following a two-step approach to deal with outliers. First, obtain a consistent estimate of α using a robust estimator with a high "breakdown value" (i.e., not strongly affected by the presence of outliers).[8] Second, use a threshold rule to identify extreme realizations of the residuals w_j (Rousseeuw and Leroy 1987). This is done by computing the distribution of the standardized residuals, or z-scores, using the estimates from the first-stage: $\hat{z}_j = \hat{w}_j / \hat{\sigma}_w$, where $\hat{\sigma}_w$ is a robust estimate of the standard deviation of the residuals $\hat{w}_j$. Formally, the set of "influential" observations or potential outliers is defined as the individuals who experienced unusually large changes over the full treatment window: $|\hat{z}_j| > \bar{z}$. A common practice is to set a threshold of $\bar{z} = 2.5$.

A limitation of robust estimators, relative to least squares, is that they rely on nonlinear regression techniques, which can be an issue when the number of parameters is large. This is particularly relevant for DiD, due to the presence of individual and time fixed effects, which leads to an incidental parameters problem and inconsistent estimates of α. To get around this problem, we transform the data into a long-difference format: we collapse the data at the plant by "before versus after treatment-period" level to eliminate the monthly variation and plant fixed effects (this is the specification proposed in Bertrand, Duflo, and Mullainathan 2004):

$$\Delta \bar{y}_j = \alpha \mathbf{1}(divest)_j + \beta + w_j. \tag{3}$$

This leads to a robust estimate of (β, α) and σ_w, as well as a direct estimate of the relevant residuals $\hat{w}_j$. Equation (3) does not require estimating a high-dimensional vector of fixed effects, and does not suffer from an incidental parameters problem. It is important to note that long-differencing is not an appropriate approach in general to address the incidental parameters problem in robust regressions. It is appropriate in our case because we use it for detection purposes. Under the null hypothesis of a homogeneous treatment effect and no outliers, robust regressions with long-differencing will yield a consistent estimate of the average treatment effect. In the case of quantile regression, this is because the median treatment effect is equal to the average treatment effect.

The outlier diagnosis strategy relies on comparing the parameter estimates, $\hat{\alpha}^{robust}$ and $\hat{\alpha}^{ols}$ from equation (3).[9] If the two estimates are very different and outliers

[7] The presence of outliers creates strong correlation across observations (in the context of this paper: utilities or plants). Outliers can thus be viewed as a form of misspecification which leads to severe heteroskedasticity. As a result, standard errors estimates will differ greatly between clustering at different levels when there are multiple outliers in terms of j.

[8] Several robust estimators are available in the literature to perform the first stage: least-absolute value regression (or median regression), M-estimators (or Huber), S-estimators (or least-trimmed regression), and MM-estimators. In STATA, the user-written package *robreg* incorporates the most common robust estimation procedures (Jann 2010). In the context of average treatment effects, the median regression and M-estimators are known to have good robustness properties.

[9] This long-difference specification is also justified because the use of long-term contracts induces a large degree of autocorrelation in prices in our data, consistent with the fact that contracts are infrequently renegotiated (Joskow 1987).

Table 1—Plant-Level Robust Analysis

	Least squares		Robust estimation			
			Q50	Bi.	S	MM
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. DiD estimates of log(*Price*)						
1(*divest*)	−0.118	−0.125	−0.038	−0.050	−0.032	−0.049
	(0.045)	(0.039)	(0.044)	(0.031)	(0.034)	(0.032)
	[0.082]					
Standard deviation (residuals)	0.190	0.259	0.214	0.216	0.212	0.212
Data	Monthly	Long-diff	Long-diff	Long-diff	Long-diff	Long-diff
Average number of matched neighbors	6.5	6.4	6.4	6.4	6.4	6.4
Year-month fixed effects	Yes	No	No	No	No	No
Plant fixed effects	Yes	No	No	No	No	No
Plant-level clusters	206					
Utility-level clusters	98					
Synthetic-level clusters		174	174	174	174	174
Divested plants	87	87	87	87	87	87
Control plants	119	87	87	87	87	87
Breakdown point			50		25	25
Gaussian efficiency				95	76	95
R^2	0.720	0.055				
Observations	38,093	174	174	174	174	174
	Divested	Not divested	Total			
Panel B. Robust statistics						
$\sum_j \mathbf{1}(\lvert z_j\rvert > 2.5)$	7	0	7			
$\sum_j z_j \cdot \mathbf{1}(\lvert z_j\rvert > 2.5)$	−30.2	0.0	−30.2			

Notes: In panel A, the dependent variable is the logarithm of the coal price. *Least squares* indicates the synthetic control matched OLS regressions. Matching is based on distance and coal type in 1997. Maximum nearest neighbors is 10. *Q50* indicates median regression. *Bi.* indicates Tukey's biweight estimator (a type of M-estimator). *S* indicates an S-estimator. *MM* indicates an MM-estimator. Standard errors are clustered at the plant level in parentheses and at the utility level in square brackets for least-squares regressions, and at the synthetic control level for robust regressions. In panel B, the robust statistics are based on the MM-estimator in column 6 of panel A.

are present, it suggests a violation of the hypothesis, either caused by the presence of an outlier, or heterogeneous treatment effects.

B. *Estimation Results*

Table 1 presents the results.[10] As in Cicala (2015), we account for observed differences across treated and control plants using a matching estimator (following Abadie, Diamond, and Hainmueller 2010 and Heckman et al. 1998). We present

[10] In the interest of space, we direct the reader to Cicala (2015) for the institutional details, description of the data, and details of the research design. We independently build our data from the original survey forms collected by the Energy Information Administration (Energy Information Administration 1990–2009). We supplement these data with power plant M&A data from Thomson Reuters (1990–2009), power plant environmental equipment data from Environmental Protection Agency (1996–2012), and incentive regulation data from Regulatory Research

the clustered standard errors in brackets below each point estimate. Column 1 of panel A shows that our matched OLS DiD estimate at the plant-month level replicates the 12 percent treatment effect in Cicala (2015).[11] Column 2 reports the DiD estimate using the long-difference regression at the plant level. This increases the treatment effect ever so slightly (from 11.8 percent to 12.5 percent).

The OLS point estimates are statistically significant at the 1 percent level when clustering at the plant level, but the p-values are greater than 0.1 when clustering at the utility-operator level. This difference is due to the presence of strong correlation between residuals belonging to the same utility-operator, which provides justification to cluster standard errors at the utility level. It is also our first hint that outliers are present in the data. In ComEd's case, the correlation is caused by their plants paying much-above-market prices. In general, negotiations at the utility level (certainly for ComEd; possibly also for other utilities given that these were vertically integrated monopolists facing rate-of-return regulation) is a potential explanation for the presence of correlated residuals, but we do not want, and neither do we need to, claim that all negotiations are at the utility level. With utility-level clustering, the 11.8 percent treatment effect is not statistically significantly different from zero, an important caveat when interpreting the results of prior literature.

Columns 3–6 present the robust regression results. We compare the estimates from four estimators: quantile (median) regression, M-estimator (Tukey's biweight estimator; a type of M-estimator), S-estimator, and MM-estimator. The robust point estimates range from -3.2 percent to -5.0 percent. We fail to reject the null hypothesis of a zero treatment effect in all cases. We obtain similar, but slightly smaller, estimates when matching on the pre-divestiture incentive regulatory status, in addition to plant location and fuel type: the robust estimates range from -0.4 percent to -2.5 percent. Online Appendix Section A.3 presents those additional results.

The fact that the robust estimates differ substantially from the OLS estimates is consistent with the presence of one or multiple outliers. Thus, we construct the distribution of robust standardized residuals (or robust z-scores) to identify extreme realizations of the residuals, and flag certain plants to be highly influential using a threshold rule.

In panel B of Table 1, we report the number and the magnitude of the statistical outliers. In calculating the standardized residuals, we use the robust estimates in column 6 (i.e., an MM-estimator with a breakdown point of 25 percent and Gaussian efficiency of 95 percent).[12] Using Rousseeuw and Leroy's (1987) standard cutoff of 2.5, we identify seven plants to be statistical outliers from the treatment group. From equation (2), we know that the impact from outliers is proportional to the long-difference residuals w_j. After normalizing by the standard deviation of these residuals σ_w, the magnitude of the bias expressed as the sum of the z-scores of the

Associates (1990–1998). We also cross-check the validity of our data to the replication data available for Cicala (2015), available at https://doi.org/10.3886/E112957V1.

[11] The matched sample is composed of 206 distinct plants, and 78 utility-operators. Our 11.8 percent estimate is very close to, but not exactly the same as the 12.4 percent in Cicala's Table 2, column 4. We can also replicate the 12.4 percent, but choose to make a minor improvement to the definition of coal type used in 1997. This has a minor effect on the matching. Details available on request.

[12] Gaussian efficiency is the efficiency of a robust estimator relative to a least-squares estimator. The notion of a breakdown point is, roughly, the smallest amount of contamination that may cause an estimator to take on arbitrarily large aberrant values (Donoho and Huber 1983).

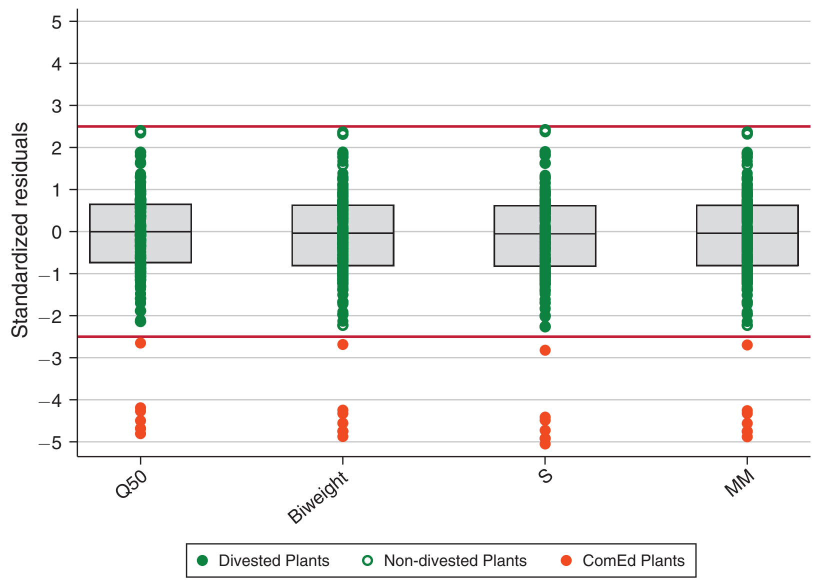

FIGURE 1. DISTRIBUTION OF PLANT-LEVEL ROBUST STATISTICS

Notes: Q50 indicates median regression (breakdown point: 50 percent). *Biweight* indicates Tukey's biweight estimator (a type of M-estimator) with Gaussian efficiency set to be 95 percent. *S* indicates an S-estimator with the breakdown point set to be 25 percent (equivalently, Gaussian efficiency to be about 76 percent). *MM* indicates an MM-estimator with the first-stage breakdown point set to be 25 percent and the second-stage Gaussian efficiency set to be 95 percent.

outlier residuals is −30.2. This sum is large and negative, and the outliers are all part of the treatment group. This implies that the outlier plants experienced long-term price changes that are on average four times the standard deviation of residual prices (i.e., $-30.2/7 = 4.3 = \bar{w}^{outliers}/\sigma_w$), well above standard levels. The outliers thus strongly pull down the OLS estimate.

The distribution of the z-scores from each robust estimator is plotted in Figure 1. The figure shows that regardless of the specification, there are seven treated plants that stand out "too much." These seven plants all belong to ComEd, a large electric utility in Illinois. Importantly, every ComEd plant in our sample is flagged as an outlier. The fact that all outlier plants belong to the same utility explains the difference between the estimated standard errors calculated by clustering at the plant versus utility level in column 1 of panel A.

Finally, it is important to mention that the results differ from the ones obtained using outlier detection approaches commonly used in economics. In applied econometrics, detecting outliers is often done by inspecting the distribution of least-squares residuals, or measuring the sensitivity of OLS parameter estimates to individual observations using "leave-one-out" regression techniques (e.g., based on the *dfbeta* statistic). Online Appendix Figure A.6c shows that this approach fails to detect the presence of outliers in our application. The leave-one-out regression detects the presence of outliers only when we remove groups of observations (online Appendix Figure A.6b), which is ex ante unknown in most applications.

In general, these detection methods are not valid since the least-squares regression line is biased in the direction of the outlier observations and are sensitive to the presence of multiple outliers, a phenomenon known as the *masking effect*. The robust-regression approach does not suffer from this problem, since a consistent estimate of α is obtained first before detecting outlier residuals.

II. Discussion

A. *ComEd's Contracting History*

That all of the robust estimators, which are also consistent estimates of the treatment effect, differ substantially from the OLS estimators, which are known to be sensitive to outliers, is important in its own right and motivates further analysis. A single utility drives at least 50 percent of the treatment effect, which is critical to know for external validity. The next question to consider is whether ComEd is an invalid outlier.

Outliers can be invalid because of data errors (unlikely here), or because of model misspecification (e.g., an omitted variable). For example, in the DiD context an outlier is invalid if it is associated with large residual changes in the outcome variable that are not caused by the policy change. In both cases, the presence of outlier residuals suggests that individuals are not drawn from a common distribution in the treated and control groups, which calls into question the main identifying assumption (e.g., common trend and support). In contrast, if the outliers are caused by heterogeneity in the treatment effect (an unusually large response of ComEd to divestiture), dropping or muting the effect of influential observations will bias the estimate, referred to as the *swamping effect* in the statistics literature. To determine which of the two possibilities is more likely, we turn to ComEd's institutional history and characteristics.

ComEd had a rather unique history in terms of its coal procurement contract (re-)negotiation. The company had signed long-term contracts, which are negotiated at the company-level, not at the plant-level, in the 1970s that turned out to be very unfavorable ex post. Specifically, ComEd purchased the vast majority of its coal under long-term contracts lasting until the 2000s with the Decker mine in Montana and the Black Butte mine in Wyoming, both owned by the same parent company (Peter Kiewit & Sons and partners). These contracts were reasonable given the energy crisis in the 1970s, but the remarkable growth of the sub-bituminous coal sector in Montana and Wyoming, which was driven by railroad deregulation under the Staggers Rail Act of 1980 and increased demand for lower sulfur coal, had led to substantially lower spot prices by the mid-1980s (Ellerman et al. 2000).

While other utilities either renegotiated or bought out their previously signed long-term contracts in the 1980s,[13] ComEd was only able to negotiate delaying the deliveries of coal and purchased some of its commitments as "reserve coal" left in the ground, with the option to have the mines extract and ship it in later years. Thus, ComEd already owned and had paid for the option to extract the reserve coal accumulated over the period 1982–1992 under a Coal Lease Purchase Agreement. As a

[13] See John N. Maclean, "Ratepayers Stand to Gain from Challenge to Edison Coal Costs," *Chicago Tribune*, May 6, 1992.

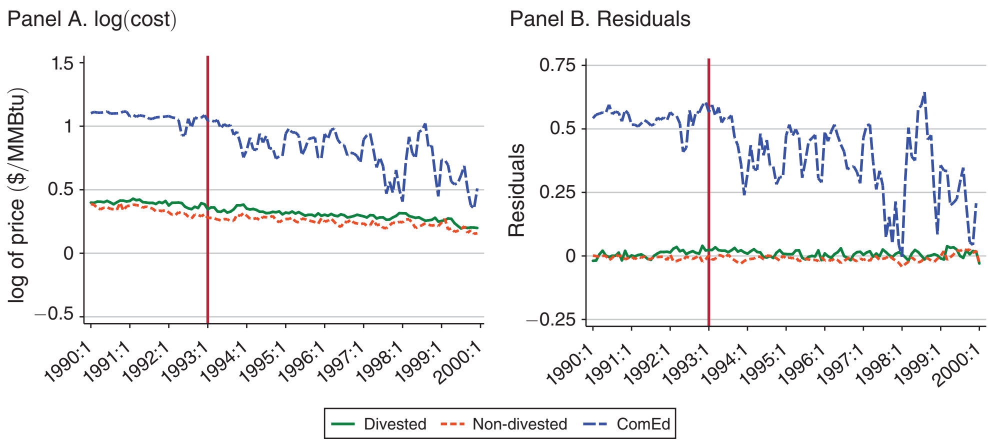

FIGURE 2. PRE-TRENDS IN DELIVERED COAL PRICES FOR COMED PLANTS

Notes: Panel A shows the weighted average of log delivered coal costs in \$/MMBtu separately for divested (excluding ComEd), non-divested, and ComEd plants. Non-divested plants receive a weight $1/m_j$ where m_j is the number of non-divested plants matched to a divested plant j. Panel B shows de-trended averages of log delivered coal costs separately for divested (excluding ComEd), non-divested, and ComEd plants. The de-trended averages for the divested (excluding ComEd) and non-divested plants are the residuals from a regression of log cost on year-month and plant fixed effects. See footnote 13 for the details of the de-trending procedure and the calculation of the ComEd residuals. The red vertical line indicates ComEd's renegotiation in late 1992, which became effective as of January 1, 1993.

result, by the early 1990s ComEd was still tied to its long-term contracts, ended up paying about twice the market price, and had accumulated substantial amounts of reserve coal.

In late 1992, however, ComEd successfully renegotiated its contracts with Decker and Black Butte, after which delivery prices went down. Figure 2 shows a pronounced decreasing trend in ComEd's fuel costs. Panel A shows the cost reductions for ComEd in the pre-period in the raw data, as well as the (much more modest) cost reductions for other treatment plants and for control plants. Panel B shows the same trends, but now for residualized fuel costs.[14] Before ComEd's contract renegotiation in late 1992, average residualized fuel prices were flat. In contrast, there was a pronounced decreasing trend between the 1992 renegotiation and the divestiture in the year 2000. A further decomposition of the data reveals that (i) prices for both the Decker and the Black Butte contract exhibited a pronounced downward trend between 1993 and 2000; (ii) the price variance under the Black Butte

[14] In order to avoid a masking effect, we residualize log costs using the sample of the data that excludes the ComEd plants. We implement this in two stages: (i) de-trend log costs over time and (ii) de-mean for each plant. In the first step, we estimate a time trend and a treatment effect *only on the subsample that excludes ComEd* (i.e., we estimate the time fixed effects and a DiD indicator in the following regression: $\log(cost)_{jt} = \gamma_t + \mathbf{1}(Divest)_{jt} + \epsilon_{jt}$). We then de-trend log costs by the estimated trend *for the entire sample* (i.e., $\hat{\epsilon}_{jt} = \log(cost)_{jt} - \hat{\gamma}_t$). In the second step, we de-mean the de-trended log costs for each plant such that for each plant, residualized log costs are centered around 0 (i.e., $\hat{\epsilon}'_{jt} = \hat{\epsilon}_{jt} - \left(1/(t^*_j - 1)\right)\sum_{t=1,\ldots,t^*_j-1}\hat{\epsilon}_{jt}$). For all plants excluding ComEd, this procedure is equivalent to obtaining the residuals from a regression of log cost on year-month and plant fixed effects. For the ComEd plants, we adjust log costs by the overall trend of the other plants and de-mean log costs by their own plant-level average. Note that the ComEd residuals are centered around zero over the entire sample period, but for data confidentiality reasons we can only report the results for the pre-period.

contract increased sharply (Decker's price variance is always small); (iii) ComEd increased delivery quantities under the Black Butte contract, stockpiling the surplus; and (iv) ComEd started buying coal on the spot market at low prices (to reduce the average price further and reduce regulatory pressure).

These empirical observations are consistent with what we know about the 1992 renegotiation. Starting in 1988, the city of Chicago, the Illinois State Attorney's Office, the state's Office of Public Counsel, and extensive media coverage pressed the issue of ComEd's high coal costs.[15] The state regulator, the Illinois Commerce Commission (Illinois' Public Utility Commission equivalent), investigated ComEd's fuel expenditures, which eventually led to the Fuel Matters Settlement in 1993. Under this agreement, ComEd had to reimburse its customers for "imprudently incurred" Western coal expenditures between 1989 and 1992. The increased regulatory surveillance set clear expectations that ComEd would not be allowed to fully pass through all of its above-market fuel costs in the future. Facing additional incentives to reduce its fuel costs going forward even as a regulated utility, ComEd renegotiated with Decker and Black Butte in 1992.

In addition, ComEd had improved its bargaining position by 1992. Two factors can explain this change. First, ComEd could have decided not to exercise the option to extract the coal reserves paid for and accumulated over the period 1982–1992 and source their coal under better market conditions, while the mines could not sell this reserve coal to a buyer other than ComEd until the expiration date (unknown to us). Second, the mines were no longer competitive in the 1990s, and would have had difficulty finding a profitable buyer in the event that ComEd had refused to increase deliveries from Black Butte and let the purchase agreement lapse. This reduced the bargaining leverage of the mines and allowed ComEd to negotiate better terms with the mines. ComEd's 1993 10-K report states that "the Company's western coal contracts and its rail contracts for delivery of the western coal were renegotiated […] to provide, among other things, for significant reductions in the delivered price of the coal over the duration of the contracts."

How did the renegotiation change pricing? Although the exact terms of the renegotiated contracts are not public, we know a few important details from press releases, the 10-K reports of ComEd and the coal suppliers, and various audits into ComEd's coal contracts by the government agencies. First, ComEd agreed to commit to have its accumulated coal reserves extracted. Effective in 1993, all reserve coal would now be delivered from the Decker mine in Montana and taken out of the Black Butte contract. ComEd would pay a low price of \$7/ton for the extraction of reserve coal (Decker Coal Company, 176 IBLA 277, 2009). Second, the rest of the coal commitments under the Black Butte contract were transferred to and set to be delivered from two unaffiliated mines in Wyoming (Peter Kiewit & Sons Incorporated 1994). The decreasing trend for Wyoming coal and the increase in price variance clearly

[15] For more details, see John Boyd, "Utility Pact Fuels Surge in CNW Coal Loadings." *Journal of Commerce*, February 22, 1993; Rob Karwath, "ICC Wants Edison Proof of Coal Cost or Customer Refund." *Chicago Tribune*, August 11, 1991; Rob Karwath, "Ruling on Coal Contracts Says Edison 'Prudent,'" *Chicago Tribune*, July 24, 1993; John N. Maclean, "Edison Will Save Two Billion Dollars on Coal," *Chicago Tribune*, 1992; John N. Maclean, "Ex-Edison Exec Defends Keeping Coal Contracts," *Chicago Tribune*, 1992; John N. Maclean, "Ratepayers Stand to Gain from Challenge to Edison Coal Costs"; and Commonwealth Edison Company (1993). For example, it was discovered that ComEd was paying \$55.62 per ton while Detroit Edison, a utility in Michigan, was paying \$30.16 for the same coal from the Decker mine in 1990 (Maclean, "Ratepayers Stand to Gain," ibid.).

suggest that the renegotiation had changed the structure and the pricing, a likely explanation is that the contract was renegotiated from a fixed-price contract to a cost-plus contract under which a fixed price applies to a minimum quantity, and any extra quantity is sold at a variable price that reflects market conditions. ComEd's increasing deliveries of Wyoming coal from 1993–1999 then explain the decreasing price trend. The price for Decker coal decreased gradually as well.[16] Third, ComEd agreed to purchase rail cars which reduced transportation cost going forward and also contributed to a lower delivery price post-1992 (although this explains a level shift more so than the decreasing trend).

This chain of events helps explain the pronounced pre-trend in Figure 2. A linear trend fitted through ComEd's residuals (allowing for serial correlation through an AR(1) process) yields a statistically significant downward trend in ComEd's residual fuel prices starting in 1993, accumulating to an average 33.6 percent residual price decline between 1993 and 2000 (also visible in panel B). The residualized trend between 1993 and 2000 is substantial although somewhat lower in magnitude than the $\sim$50 percent price decline in the raw data in panel A, since procurement costs for coal decreased in general in the 1990s.

Despite this pre-trend for ComEd, the parallel trends assumption in the full sample cannot be rejected, and there is no clear visual evidence that indicates misspecification of the DiD model. This suggests that an outlier realization of w_j does not necessarily lead to a rejection of the standard pre-trend tests in a DiD framework. See online Appendix Section A.1 for a discussion of the impact of outliers on pre-trend tests.

We cannot rule out the possibility that some of the price decline observed after 2000 is related to divestiture. In December 1999, as part of the negotiations during the sale of ComEd's assets, ComEd made a lump-sum payment of $350 million to Midwest Generation to compensate them for the above-market portion of the coal purchase commitments with Decker. Midwest then bought out the Decker contract and coal deliveries stopped in January 2003. Combined with the end of the Black Butte contract, this led to a further decrease in the delivery price post-2002.

However, there are various reasons to believe that the post-2000 price decline would have been observed even absent of divestiture. First, the Black Butte contract ended in 2000 after a period of frontloading deliveries by ComEd (initiated in 1993). Our understanding is that this was enabled by the 1992 renegotiation as two unaffiliated mines took over production on behalf of Black Butte, allowing ComEd to fulfill its obligations earlier than the original 2007 end date. After 2000, ComEd would have signed new contracts at market prices even absent divestiture, once it had burned through its stockpiles. Second, in December 1999, ComEd had substantial commitments of reserve coal left. ComEd had started to have some of these reserves extracted at lower prices in the late 1990s (leading to a decreasing price for Decker coal), and would have likely continued ordering the remaining cheap reserve coal in increasing amounts in the 2000s, thereby continuing the downward pre-trend

[16]Explanations include (i) an increasing share of reserve coal under a two-tiered pricing system for Decker coal: the average price depends on the fraction of coal delivered under the lower "reserve price" versus the higher base price negotiated in 1970s; (ii) a change in the contractual pricing agreement for new post-1992 commitments with Decker, but we could not find any direct information about this.

observed in the data. Finally, the Decker buy-out could have happened even in the counterfactual environment without deregulation. For example, continued regulatory pressure could have triggered another round of renegotiation. In addition, increased M&A activities in the industry affected not only divested companies, but also utilities located in regulated states (control group). Therefore, ComEd could have merged with a larger entity, and bought out the Decker contract.

In sum, it is possible that divestiture caused (in part) the Decker buy-out which led to cost savings. However, it is also clear that the 1992 renegotiation affected ComEd's procurement strategy over the entire period (including beyond 2000). This separate event occurred prior to and was unrelated to divestiture. The first-ever hearing about restructuring in Illinois happened in 1995, well after ComEd had renegotiated its contracts (Regulatory Research Associates 1996). Therefore, the contract renegotiation confounds the effect of divestiture, and it is impossible to separate the two without a comparable control. Unfortunately, no non-divested plants experienced a similar contracting and renegotiation history as ComEd, making it difficult to have a credible counterfactual reference for the change in ComEd's costs in the absence of divestiture. Such a counterfactual should incorporate the gradual and long-lasting decline in prices beginning in 1993 as a result of the renegotiation in 1992.[17] The lack of suitable control plants is the central problem caused by outliers in DiD settings.

B. *Average Treatment Effect Accounting for Outlier Plants*

Having established that the standard DiD matching estimators suffer from misspecification caused by ComEd's 1992 contract renegotiation, we present three options for estimating the average treatment effect of divestiture on coal procurement costs using a least-squares framework. The first solution is to drop ComEd from our sample. The two other solutions are to model a counterfactual by accounting for ComEd's pre-deregulation procurement history (i.e., steadily declining coal prices following the 1992 renegotiation) through a flexible ComEd-specific time trend in the regression specifications. We thus include a counterfactual for the entire sample period using the pre-trend that we observe for 1993–2000; our best-available evidence. We suggest two ways of implementing this; both motivated by Figure 2.

The latter solutions still allow ComEd plants to contribute to the treatment effect at the time of deregulation (relative to a decreasing time trend). In the absence of deregulation, the decreasing cost trend for ComEd plants would likely have continued as per the renegotiated contract terms in 1992, so the counterfactual should include a decreasing time trend starting in 1993 and running through the end of our sample. In the second solution, we include a piecewise linear time trend for ComEd that is flat during 1990–1992 but can increase or decrease linearly from January 1993 through the end of our sample. This second option corresponds to the following specification:

$$\text{(4)} \quad \log\left(price\right)_{jt} = \alpha \mathbf{1}\left(divest\right)_{jt} + \beta \mathbf{1}\left(ComEd\right) \cdot \max\left[time_t - 35, 1\right] + \gamma_j + \delta_t + \epsilon_{jt}$$

[17] John Boyd, "Utility Pact Fuels Surge in CNW Coal Loadings." *Journal of Commerce.*

TABLE 2—DID ESTIMATES OF LOG(PRICE) ACCOUNTING FOR COMED CONTRACT RENEGOTIATION

	No trend, drop ComEd (1)	ComEd-specific trend (2)	De-trended with ComEd projected post-trend (3)
$\mathbf{1}(divest)$	−0.062	−0.050	−0.064
	[0.044]	[0.045]	[0.047]
$\mathbf{1}(ComEd) \cdot \max[time - 35, 1]$		−0.006	
		[0.000]	
Drop ComEd	Yes	No	No
Average number of matched neighbors	6.3	6.5	6.5
Year-month fixed effects	Yes	Yes	Yes
Plant fixed effects	Yes	Yes	Yes
Plant-level clusters	184	206	206
Utility-level clusters	93	98	98
Divested plants	80	87	87
Control plants	104	119	119
R^2	0.777	0.778	0.768
Observations	34,145	38,093	38,093

Notes: Dependent variable is the logarithm of the coal price. Matching is based on distance and coal type in 1997. Maximum nearest neighbors is 10. Standard errors are clustered at the utility level in square brackets. A plant's parent utility is determined based on plant ownership as of 1997. In column 3, dependent variable is de-trended using ComEd's estimated pre-trend and projected onto the post-period.

where α is the DiD coefficient of interest. The variable $time_t$ is a year-month index that takes on the value 1 in January 1990 (the first month in our sample), 2 in February 1990, etc. In the third solution, we first estimate a piece-wise linear pre-trend for ComEd during the pre-deregulation period 1990–2000. We then project this pre-trend onto the post-period 2000–2010. We then de-trend the fuel cost variable and proceed with estimating equation (1) as before.

Table 2 shows the estimation results for these various options. The treatment effect of divestiture on coal procurement cost is between −5 percent to −6 percent, much smaller in absolute value than the −12 percent estimate when we do not account for outliers (Table 1, column 1). Also note that, after accounting for the outlier, the utility-level clustered standard errors become similar to the plant-level clustered standard errors in Table 1. Using all three options, we find that the effect of divestiture decreases substantially and becomes statistically insignificant (although recall that even the 12 percent treatment effect is not significant using utility-level clustering). Finally, online Appendix Section A.2 shows that the robust regression results with a ComEd time trend as in equation (4) are similar to but even smaller than the estimates in Table 1.

In column 2, we introduce a time trend for ComEd. As detailed in online Appendix Section A.4, we run a set of regressions in which we include a time trend for all other utilities one at a time: the resulting coefficients are closely centered around −12 percent except when the trend is added for ComEd. We then run a single regression with time trends for all utilities (including ComEd) and confirm that the result remains close to −5 percent. Taken together, these findings suggest that there are no other utilities that influence the estimation by nearly as much as ComEd.

In a supplementary analysis in online Appendix Section A.5, we control for the fact that ComEd's divestiture coincided with the implementation of the Acid Raid Program, a sulfur dioxide emissions permit trading program. We provide evidence that, absent this environmental regulation, the already-low average treatment effect would have been even lower (between −2 percent and −4 percent).

In summary, we find that after accounting for the presence of outliers, the average treatment effect varies between −5 percent and −6 percent, and is not statistically different from zero. These estimates are 50 percent smaller in magnitude than the outlier-driven matched OLS estimator that does not correct for the misspecification caused by ComEd's 1992 contract renegotiation.

III. Conclusion

This paper revisits an earlier result on the effect of US electricity restructuring on fuel procurement costs. In contrast to the large and significant estimates in Cicala (2015), we find that the treatment effect is about half as large and is not statistically significant. To arrive at this estimate, we use methodologies from the robust estimation literature and find that the previously estimated large effects are mainly driven by seven plants owned by the Illinois electric utility ComEd that experienced price declines in the order of four times the standard deviation of contractual coal delivery prices. An important driver of this decline is a contract renegotiation that occurred prior to and independent of divestiture, hence confounding the earlier treatment effect estimates.

Our paper illustrates the usefulness of robust methodologies in the context of estimation of treatment effects. One takeaway from our work is that it would be good practice for empirical researchers to complement their usual estimates with estimates using robust methodologies, especially because standard specification tests like balance-on-covariate tests and parallel pre-trend tests can fail to pick up outlier bias even if it is present. These robust methodologies are easy to use and are readily available.

Finally, our paper shows the importance of accounting for potential heterogeneous effects and understanding the underlying mechanisms behind the treatment. One such driver of heterogeneous effects is exposure to incentive regulation. Knittel (2002) finds that plants subjected to modified fuel adjustment clauses and heat rate efficiency programs experienced an increase in fuel efficiency. Incentive regulation may thus potentially limit the scope through which divestiture can yield further cost reductions. In ongoing work, we explore the role of previous incentive regulation, bargaining power, and contract disadvantages in explaining the effect of deregulation on fuel procurement (Han et al. 2020).

REFERENCES

Abadie, Alberto, Alexis Diamond, and Jens Hainmueller. 2010. "Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program." *Journal of the American Statistical Association* 105 (490): 493–505.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" *Quarterly Journal of Economics* 119 (1): 249–75.

Chan, H. Ron, Harrison G. Fell, Ian Lange, and Shanjun Li. 2017. "Efficiency and Environmental Impacts of Electricity Restructuring on Coal-Fired Power Plants." *Journal of Environmental Economics and Management* 81: 1–18.

Cicala, Steve. 2015. "Replication Data for: When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation." American Economic Association [publisher], 2015. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2019-10-12. https://doi.org/10.3886/E112957V1 (accessed September 5, 2016).

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

Commonwealth Edison Company. 1993. "1993 10-K Form." http://www.getfilings.com.

Conley, Timothy G., and Christopher R. Taber. 2011. "Inference with 'Difference in Differences' with a Small Number of Policy Changes." *Review of Economics and Statistics* 93 (1): 113–25.

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

Decker Coal Company. 2009. *176 IBLA 277*. United States Department of the Interior, Office of Hearings and Appeals. https://www.oha.doi.gov/IBLA/Ibladecisions/176IBLA/176IBLA277.pdf.

Donoho, David L., and Peter J. Huber. 1983. "The Notion of Breakdown Point." In *A Festschrift for Erich L. Lehmann*, edited by Peter J. Bickel, Kjell A. Doksum, and J. L. Hodges, 157–84. Belmont, CA: Wadsworth.

Ellerman, A. Denny, Paul L. Joskow, Richard Schmalensee, Juan-Pablo Montero, and Elizabeth M. Bailey. 2000. *Markets for Clean Air: The U.S. Acid Rain Program*. Cambridge, UK: Cambridge University Press.

Energy Information Administration. 1990–2009. "Data on Power Plant Divestitures, Fuel Expenditure, Operation and Characteristics from 'Electric Power Monthly Tables' and Survey Forms EIA-423, FERC-423, EIA-759, EIA-767, EIA-860, EIA-861, EIA-876, EIA-906, and EIA-923." https://www.eia.gov/Survey/ (accessed September 5, 2016).

Environmental Protection Agency. 1996–2012. "Data on Power Plant Divestitures and Locations from Air Markets Program Data and eGRID." https://ampd.epa.gov/ampd/ and https://www.epa.gov/egrid/download-data (accessed January 13, 2016).

Fabrizio, Kira R., Nancy L. Rose, and Catherine D. Wolfram. 2007. "Do Markets Reduce Costs? Assessing the Impact of Regulatory Restructuring on U.S. Electric Generation Efficiency." *American Economic Review* 97 (4): 1250–77.

Fowlie, Meredith. 2010. "Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement." *American Economic Review* 100 (3): 837–69.

Han, Jin Soo, Jean-François Houde, Arthur A. van Benthem, and Jose Miguel Abito. 2020. "Heterogeneous Effects of Deregulation on Fuel Procurement Costs: Evidence from U.S. Electricity Restructuring." Unpublished.

Han, Jin Soo, Jean-François Houde, Arthur A. van Benthem, and Jose Miguel Abito. 2021. "Replication Data for: When Does Regulation Distort Costs? Lessons from Fuel Procurement in US Electricity Generation: Comment." American Economic Association [publisher], Inter-university Consortium for Political and Social Research [distributor]. https://doi.org/10.3886/E124101V1.

Heckman, James, Hidehiko Ichimura, Jeffrey Smith, and Petra Todd. 1998. "Characterizing Selection Bias Using Experimental Data." *Econometrica* 66 (5): 1017–98.

Hubert, Mia, Peter J. Rousseeuw, and Stefan Van Aelst. 2008. "High-Breakdown Robust Multivariate Methods." *Statistical Science* 23 (1): 92–119.

Jann, Ben. 2010. "ROBREG: Stata Module Providing Robust Regression Estimators." Statistical Software Components S457114. Boston College, Department of Economics.

Jha, Akshaya. 2019. "Dynamic Regulatory Distortions: Coal Procurement at U.S. Power Plants." Unpublished.

Joskow, Paul L. 1987. "Contract Duration and Relationship-Specific Investments: Empirical Evidence from Coal Markets." *American Economic Review* 77 (1): 168–85.

Knittel, Christopher R. 2002. "Alternative Regulatory Methods and Firm Efficiency: Stochastic Frontier Evidence from the U.S. Electricity Industry." *Review of Economics and Statistics* 84 (3): 530–40.

Laffont, Jean-Jacques, and Jean Tirole. 1993. *A Theory of Incentives in Procurement and Regulation*. Cambridge, MA: MIT Press.

Malik, Kabir, Maureen Cropper, Alexander Limonov, and Anoop Singh. 2011. "Estimating the Impact of Restructuring on Electricity Generation Efficiency: The Case of the Indian Thermal Power Sector." NBER Working Paper 17383.

Newbery, David M., and Michael G. Pollitt. 1997. "The Restructuring and Privatisation of Britain's CEGB: Was It Worth It?" *Journal of Industrial Economics* 45 (3): 269–303.

Ontiveros, Darwin Ugarte, Gustavo Canavire-Bacarreza, and Luis Castro Peñarrieta. 2017. "Outliers in Semi-Parametric Estimation of Treatment Effects." Development Research Working Paper 06/2017.

Peter Kiewit & Sons Incorporated. 1994. "1994 10-K Form." http://www.getfilings.com.

Regulatory Research Associates. 1990–1998. "Data on Power Plant Incentive Regulation." Unpublished data (accessed September 5, 2016).

Regulatory Research Associates. 1996. *Regulatory Focus: Illinois Regulation Annual Review*. Jersey City, NJ: Proprietary Annual Reports.

Rousseeuw, Peter J., and Annick M. Leroy. 1987. *Robust Regression and Outlier Detection*. Hoboken, NJ: Wiley & Sons.

Sølvsten, Mikkel. 2020. "Robust Estimation with Many Instruments." *Journal of Econometrics* 214 (2): 495–512.

Thomson Reuters. 1990–2009. "Data on Power Plant M&As." Unpublished data (accessed September 5, 2016).

United States Census Bureau. 2016. "Data on Cartographic Boundary (GIS Shapefile)." https://www.census.gov/geographies/mapping-files/time-series/geo/carto-boundary-file.html (accessed September 5, 2016).

Wolfram, Catherine. 1999. "Measuring Duopoly Power in the British Electricity Spot Market." *American Economic Review* 89 (4): 805–26.

Young, Alwyn. 2020. "Consistency without Inference: Instrumental Variables in Practical Application." Unpublished.