A Online Appendix

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

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

A.1 Balance of Covariates and Test for Pre-trends

Empirical results

We show two standard tests to assess the validity of a DiD estimator: a balance-of-covariates test and a test for equal pre-trends. Table A.1 presents the balance-of-covariates test. The table illustrates that divested and non-divested plants are well-balanced in the matched sample except for the delivered coal prices. The fact that plants in the treated sample paid 7-12% higher coal prices prior to divestiture is somewhat surprising, since matching restricts the econometrician to compare plants based on coal type and geography.[1] This difference is caused by an outlier contract, but is not necessarily inconsistent with the DiD assumption since plant fixed effects are included in the regression model.

Next, we turn to a standard pre-trend test. This is the most commonly used empirical test to assess the validity of a DiD research design. Figure A.1, Panel (a) reveals no visual evidence of differential trends between the treatment and control groups. Panel (b) shows that average prices for treated plants were higher than for control plants. In Table A.2, we test whether the two groups have differential linear trends before 1997. The relevant coefficients on $time_t * \mathbb{1}(divest)_j$ suggest that there is a slight negative pre-trend in the full sample of -0.2 percentage points per month, but the slope coefficient is not statistically significant (t-statistic: -0.5). Thus we cannot reject the null hypothesis of equal pre-trends using the standard pre-trend test. Hence, neither the balance-of-covariates test nor the pre-trend test suggests a problem with the DiD estimator, and therefore do not pick up the outlier bias in this application.

Sample Window Length, Pre-Trends and Bias

It is important to note that we identify several important outliers in the treatment group *despite* the fact that the parallel trends assumption cannot be rejected and that there is no clear visual evidence that indicates misspecification of the DiD model. This suggests that an outlier realization does not necessarily lead to a rejection of the standard pre-trend tests in a DiD framework (Angrist and Krueger, 1999). Figure A.2 shows a graphical example to convey the intuition for this. Consider a panel in which one or more cross-sectional units (individuals) in the treatment group (labeled T2) exhibit a steady long-term time trend that differs from the trend for the other treatment observations (labeled T1) and the control group (labeled C). Group T2 violates the DiD assumption of common trends. Neither T1 nor C

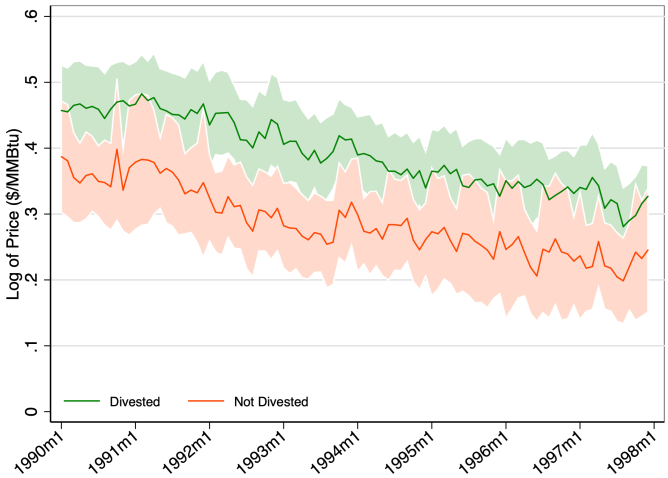
[1] Energy Information Administration (1990-2009) and Environmental Protection Agency (1996-2012) provide power plant location data. For a small number of plants missing coordinate information, we supplement the location data with the GIS data from United States Census Bureau (2016).

Table A.1: Summary Statistics for Divested and Non-Divested Plants

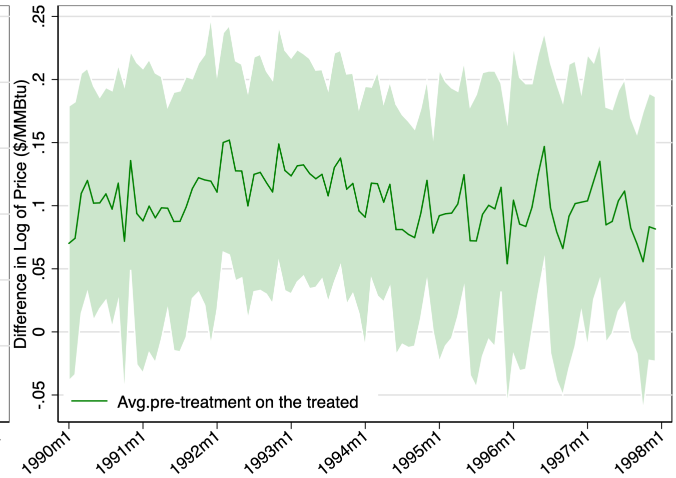
	(1)	(2)	(3)
	Divested	Not divested	Difference
Panel A: Time-invariant statistics as of 1997			
Plant vintage	1960.23	1960.15	0.08
	[11.48]	[15.27]	(2.85)
Percent scrubbers installed	0.24	0.26	-0.01
	[0.43]	[0.44]	(0.08)
Panel B: Monthly average statistics between 1990 and 1997			
Price ($/MMBtu)	1.54	1.37	0.17
	[0.47]	[0.31]	(0.06)
Log(price)	0.39	0.29	0.10
	[0.29]	[0.24]	(0.04)
Millions MMBtu delivered	3.61	3.15	0.46
	[3.55]	[3.23]	(0.57)
Percent spot market	0.26	0.28	-0.02
	[0.34]	[0.37]	(0.06)
Percent in-state	0.44	0.43	0.01
	[0.46]	[0.45]	(0.08)
Percent bituminous	0.78	0.79	-0.02
	[0.41]	[0.39]	(0.07)
Percent sub-bituminous	0.15	0.13	0.02
	[0.35]	[0.32]	(0.05)
Heat content (MMBtu/ton)	23.07	22.81	0.25
	[3.99]	[3.99]	(0.73)
Sulfur content (lbs/MMBtu)	1.42	1.60	-0.18
	[0.83]	[0.96]	(0.14)
Ash content (lbs/MMBtu)	9.81	10.05	-0.24
	[5.87]	[6.89]	(1.08)
Distance to mine (mi.)	278.21	244.56	33.64
	[304.39]	[273.55]	(44.38)
Annual capacity (MW)	973.79	867.13	106.65
	[741.41]	[702.32]	(130.92)
Annual capacity factor	0.50	0.49	0.01
	[0.19]	[0.22]	(0.04)
Plants	87	106	193

Notes: Panel A contains time-invariant statistics as of 1997. Panel B contains monthly averages between 1990 and 1997. Matching is based on distance and coal type in 1997. Maximum nearest neighbors is 10. The sample period is January 1990-December 1997. Standard deviations are in brackets for Columns (1) and (2). Standard errors are clustered at the plant level in parentheses for Column (3).

Figure A.1: Pre-Trends in Delivered Coal Prices



(a) Pre-Trends by Treatment Status



(b) Price Difference

Notes: Panel (a) shows the weighted average of log delivered coal costs in $/MMBtu for divested and non-divested plants. Panel (b) shows the difference in the weighted average of log delivered coal costs between divested and non-divested plants. In both panels, non-divested plants receive a weight $\frac{1}{m_j}$ where m_j is the number of non-divested plants matched to a divested plant j. This weighting structure is equivalent to creating a synthetic control for each divested plant. Matching is based on distance and coal type in 1997. Maximum nearest neighbors is 10. Shaded regions indicate 95% confidence intervals based on the standard errors from a regression of log cost on year-month fixed effects for treatment vs. control plants ($\log(cost)_{jt} = \gamma_t + \gamma_t \cdot 1(Divest)_j + \epsilon_{jt}$) clustered at the plant level.

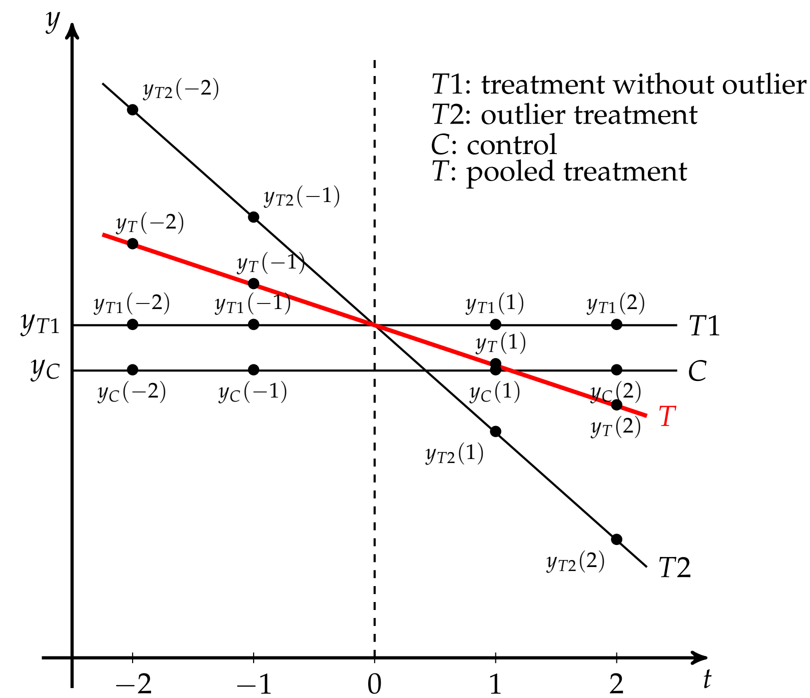
Table A.2: Hypothesis Tests on Linear Pre-Trends

	(1)
Time	-0.0016
	(0.0002)
	[0.0003]
Time·$\mathbb{1}$(Divest)	-0.0002
	(0.0003)
	[0.0004]
Avg. # of Matched Neighbors	6.3
Plant FE	Yes
Plant-Level Clusters	193
Utility-Level Clusters	89
Divested Plants	87
Control Plants	106
R^2	0.830
Observations	16,523

Notes: Dependent variable is the logarithm of the coal price. We estimate the specification $y_{jt} = \delta_0 \cdot t + \delta_1 \cdot t \cdot \mathbb{1}(divest)_j + \gamma_j + \epsilon_{jt}$, where $\mathbb{1}(divest)_j$ is a binary indicator for whether the plant was ever divested. The data is restricted to the pre-treatment period: January 1990 to December 1997. For control plants, the pre-treatment period is defined based on its matched divested plant. Matching is based on distance and coal type based in 1997. Maximum nearest neighbors is 10. Standard errors are clustered at the plant level in parentheses and at the utility level in square brackets.

exhibits a trend, so an unbiased DiD estimate based on C and T1 only would equal zero. For now, in this example, we will assume that the trend in T2 is not strong enough to reject the common pre-trend test between control observations C and the combined set of treatment observations T. The presence of the trend for group T2 will, however, affect the treatment coefficient. To see this, first consider the time window (-1,1). The slopes of the pre-trends for groups C and T are 0 and $y_{T1}(-1) - y_T(-1) = \hat{\alpha}$.[2] The DiD treatment estimate is $[y_T(1) - y_T(-1)] - [y_C(1) - y_C(-1)] = \hat{\beta} < 0$. Hence, the outlier group T2 biases the DiD estimate but this would go undetected by a pre-trend test.

Figure A.2: Outlier Bias in a Difference-in-Differences Estimator



Notes: T1 refers to the treatment observations without outliers. T2 is the group of outlier treatment observations. T represents the combined treatment group. C refers to the control group.

When the time window is expanded to (-2,2), the comparison between the pre-trend slopes for groups C and T remains unchanged: $\frac{y_{T1}(-2) - y_T(-2)}{2} = \hat{\alpha}$. The DiD treatment estimate, however, grows to $[\frac{y_T(1)+y_T(2)}{2} - \frac{y_T(-1)+y_T(-2)}{2}] - [\frac{y_C(1)+y_C(2)}{2} - \frac{y_C(-1)+y_C(-2)}{2}] = \frac{3}{2}\hat{\beta} < 0$. It is straightforward to show that, as the window expands to (-N,N), the estimate becomes $\frac{1+N}{2}\hat{\beta}$ (i.e., arbitrarily large as $N \to \infty$) while the common pre-trend test will be unaffected. Hence, an outlier observation with a long-term deviating pre-trend can introduce substantial bias in the average treatment effect estimate, even if the slope of the pre-trend is small.

We now show a numerical example of an outlier that causes a bias in the estimation but the standard pre-trend tests fail to detect any statistically significant differential trend. For each Monte-Carlo simulation, we first draw 100 control individuals, 99 treated individuals and 1 outlier (labeled as a treatment

[2]That group T has a negative slope of $\hat{\alpha}$ rather than the zero slope of the non-outlier treatment observations T1 illustrates the masking effect discussed in the main paper. In an attempt to fit the outlier(s) T2, least-squares regression fits a decreasing common trend for group T, which subsequently understates the residuals for the outliers relative to robust estimation (which would yield a slope close to zero and larger residuals in absolute value).

observation) with the following data generating process:

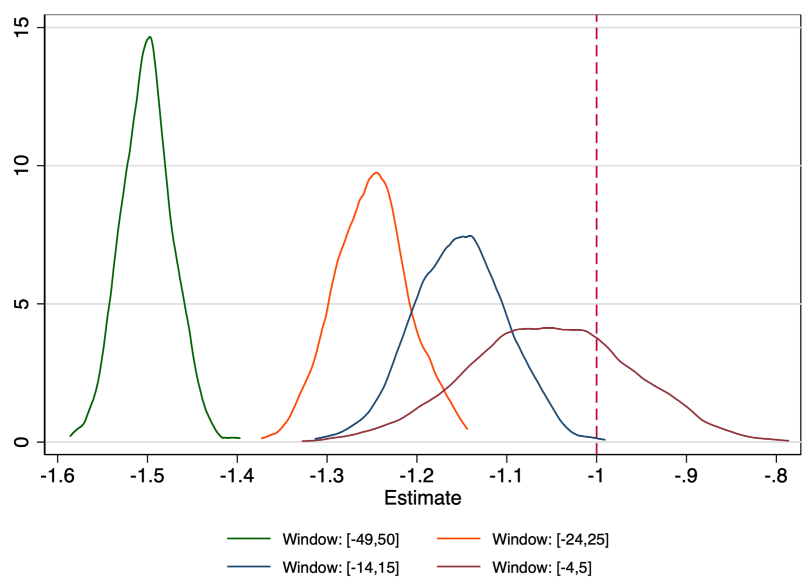
$$y_{jt} = \alpha d_{jt} + \gamma_0 + \gamma_1 \cdot t + \eta_{jt}$$

where j denotes an observation and t ranges from -49 to 50, d_{jt} equals 1 for $t > 0$ (the treatment occurs at time $t = 1$), the treatment effect $\alpha = -1$, the time trend parameter $\gamma_1 = -1$ if j is an outlier and 0 otherwise, and η_{jt} is drawn from a standard normal distribution. Then, we implement a DiD estimation and conduct a pre-trend test by varying the size of the symmetric time window around the treatment date. We choose time windows of [-4,5], [-14,15], [-24,25], and [-49,50]. We repeat the above simulation 1,000 times.

In theory, the above data generating process should yield a DiD estimate of $\alpha = -1$ in the absence of the outlier. However, the outlier biases the DiD coefficient downwards due to its declining trend over time, and the magnitude of the bias is increasing in the time window. Nonetheless, the mean differential trend for the treated group would be -0.01 (i.e., $\alpha/100$) regardless of the choice of the time window.

Figure A.3 shows the distribution of the DiD estimate from the simulations for different time windows. The bias is clearly increasing in the window size with the magnitude of the bias equal to about a half of the treatment effect as the window expands to [-49,50].

Figure A.3: Kernel Density of Simulated DiD Estimates



Notes: Epanechnikov kernel function is used to plot the kernel density. The red dashed vertical line indicates the true treatment effect, α, of the data generating process.

Table A.3 reports the average DiD estimate $\hat{\alpha}$ across the 1,000 simulations, the estimated slope of the pre-trend, and the fraction of simulations with statistically significant pre-trend tests. The average treatment effect moves further away from the true value $\alpha = -1$ as the window size grows. The linear pre-trend for the treated group is around -0.01 for all of the window sizes. For the pre-trend test, a t-test is implemented on the coefficient of a linear time trend for the treatment group. Formally, we test $H_0 : \gamma_1 = 0$ where γ_1 is the parameter from the specification $y_{jt} = \gamma_0 \cdot t + \gamma_1 \cdot t \cdot \mathbb{1}(ever\ treated)_j + \eta_{jt}$ for

$t < 51$. The pre-trend test flags less than 8% of the simulations as having a differential trend between the treated and the control group.

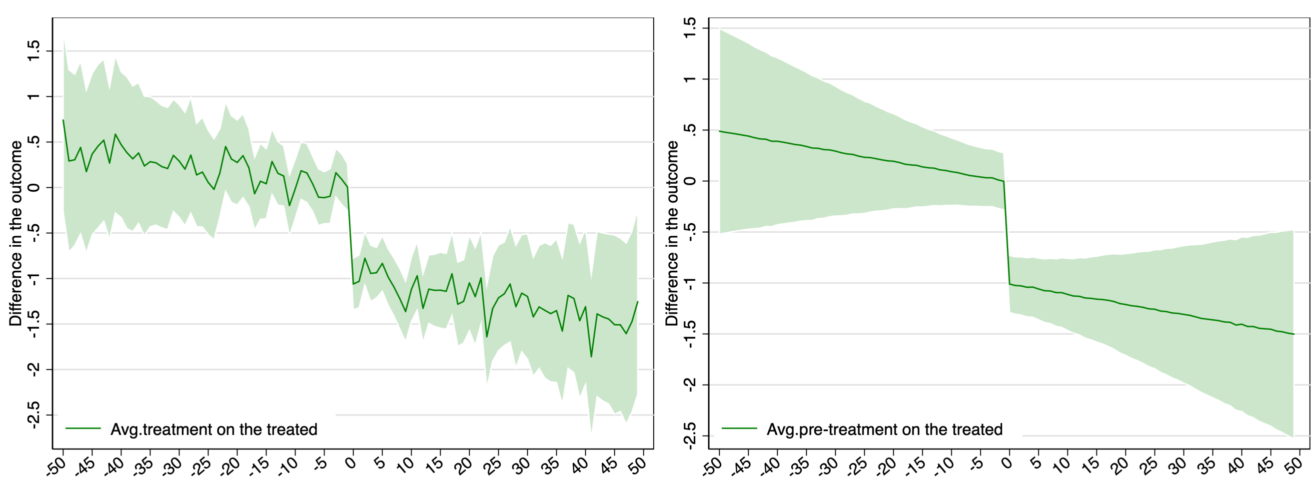
Table A.3: Simulation Statistics

Window size	$[-4, 5]$	$[-14, 15]$	$[-24, 25]$	$[-49, 50]$
Avg. DiD estimate	-1.052	-1.152	-1.250	-1.500
Avg. pre-trend	-0.010	-0.010	-0.010	-0.010
Frac. of sig. pre-trend tests	0.073	0.078	0.029	0.000

Notes: Avg. DiD estimate" denotes the average DiD estimate $\hat{\alpha}$ across the simulations. The true parameter $\alpha = -1$. "Avg. pre-trend" denotes the average $\hat{\gamma}_1$ across the simulations. "Frac. of sig. pre-trend" denotes the fraction of the simulations with p-values < 0.1 for $\hat{\gamma}_1$, where $\hat{\gamma}_1$ is the coefficient from the specification $y_{jt} = \gamma_0 \cdot t + \gamma_1 \cdot t \cdot \mathbb{1}(\textit{ever treated})_j + \eta_{jt}$ for $t < 51$.

Figure A.4 plots the event-study graphs from these simulations. The figure on the left shows a particular realization; the figure on the right shows the average over 1,000 simulations. While the pre-trend and the true treatment effect of -1 are clearly visible in the figure on the right, the single realization on the left is not quite as revealing. Based on a pre-trend test and the event-study graph, a researcher might conclude that the equal pre-trends assumption cannot be rejected and that there is no compelling visual evidence against a pre vs. post DiD analysis (which would result in an inflated treatment effect estimate). In summary, pre-trend tests and event-study figures can hide important specification problems; our proposed outlier detection method is a much more robust procedure to detect such problems.

Figure A.4: Event Study Simulation



Notes: Graphical analysis of a DiD estimator. On the left, the first simulation is reported. On the right, the average of 1,000 simulations is reported. Shaded region indicates 95% confidence interval clustered at the j level.

To establish formally why one can still have biased estimates of the average treatment effect due to outlier observations despite no apparent pre-trends, we adopt a similar specification as in our Monte Carlo simulation. Let T be the number of periods and assume that there are equal periods pre and post-treatment. Let t^* be the date of treatment. Assume that there are, respectively, J_I and J_O non-outlier and outlier units in the treatment group, and let J_C be the number of control units. Assume that there are no outlier units in the control group. Finally, we order the units by first having J_O, then J_I, and lastly J_C.

Let the true model be:

$$y_{jt} = \alpha \cdot [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{t \geq t^*\}] + \gamma_1 \cdot [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{j \in \text{outlier}\} \times t] + \epsilon_{jt}. \tag{A.1}$$

Suppose we do not take the outliers into account and estimate:

$$y_{jt} = \alpha \cdot [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{t \geq t^*\}] + e_{jt} \tag{A.2}$$

where $e_{it} = \gamma_1 [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{j \in \text{outlier}\} \times t] + \epsilon_{jt}$. The estimate for α based on Equation (A.2) is given by:

$$\hat{\alpha} = \frac{\sum_{t=1}^{T} \sum_{j=1}^{J_O+J_I+J_C} [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{t \geq t^*\}]\, y_{jt}}{\sum_{t=1}^{T} \sum_{j=1}^{J_O+J_I+J_C} [\mathbb{I}\{j \in \text{treatment group}\}\mathbb{I}\{t \geq t^*\}]^2} \tag{A.3}$$

Simplifying $\hat{a}$ gives:

$$\begin{aligned}
\hat{\alpha} &= \frac{\sum_{t=t^*}^{T} \sum_{j=1}^{J_O+J_I+J_C} [\mathbb{I}\{j \in \text{treatment group}\}]\, y_{jt}}{\sum_{t=t^*}^{T} \sum_{j=1}^{J_O+J_I+J_C} [\mathbb{I}\{j \in \text{treatment group}\}]^2} \\
&= \frac{\sum_{t=t^*}^{T} \sum_{j=1}^{J_O+J_I} y_{jt}}{\frac{T}{2}(J_O + J_I)} \\
&= \frac{\sum_{t=t^*}^{T} \sum_{j=1}^{J_I+J_O} \left[\alpha + \gamma_1 \mathbb{I}\{j \in \text{outlier}\} \times t + \epsilon_{jt}\right]}{\frac{T}{2}(J_O + J_I)} \\
&= \alpha + \gamma_1 \frac{\sum_{t=t^*}^{T} \sum_{j=1}^{J_O} t}{\frac{T}{2}(J_O + J_I)} + \frac{2 \sum_{t=t^*}^{T} \left(\frac{1}{J_O+J_I}\right) \sum_{j=1}^{J_O+J_I} \epsilon_{jt}}{T} \\
&= \alpha + \gamma_1 \frac{J_O \sum_{t=t^*}^{T} t}{\frac{T}{2}(J_O + J_I)} + o_p(1) \\
&= \alpha + \gamma_1 \frac{2J_O}{T(J_O + J_I)} \frac{(T - t^* + 1)(t^* + T)}{2} + o_p(1) \\
&= \alpha + \gamma_1 \cdot \frac{J_O(T - t^* + 1)}{J_O + J_I} \left(\frac{t^*}{T} + 1\right) + o_p(1).
\end{aligned} \tag{A.4}$$

Notice that $\hat{\alpha}$ will only be consistent if $\frac{J_I}{J_O}$ grows faster than T. In particular, if $\frac{J_I}{J_O}$ and T grow at the same rate, then $\hat{\alpha} \to^p \alpha + \gamma_1$; if T grows faster than $\frac{J_I}{J_O}$, the asymptotic bias goes to infinity. This is consistent with our Monte Carlo simulations (and the ComEd application in the main text) where the bias increases as we increase the size of the symmetric time window while keeping $\frac{J_I}{J_O}$ fixed.

Since we want our estimate of the treatment effect to be identified by non-outlier units, the number of non-outlier units relative to outlier units ($\frac{J_I}{J_O}$) should grow sufficiently fast. Specifically, we need this ratio to grow faster than the number of times we see each individual unit (T), otherwise outlier units will play an influential role in identifying the treatment effect.

Suppose that T grows faster $\frac{J_I}{J_O}$ hence our estimate of the average treatment effect is biased. Will a pre-trend test pick up the bias? That is, if we estimate the regression

$$y_{jt} = \gamma \cdot [\mathbb{I}\{j \in \text{treatment group}\} \times t] + e_{jt} \tag{A.5}$$

for $t < t^*$, will we get an estimate $\hat{\gamma}$ that converges to $\gamma_1 \neq 0$ (see Equation A.1)? If this is the case, then the test statistic using the biased estimate will not be centered at the true $\gamma = 0$ (no pre-trend).

The estimator for γ is given by:

$$\begin{aligned}
\hat{\gamma} &= \frac{\sum_{t=1}^{t^*-1} \sum_{j=1}^{J_O+J_I+J_C} \left[\mathbb{I}\{j \in \text{treatment group}\} \times t\right] y_{jt}}{\sum_{t=1}^{t^*-1} \sum_{j=1}^{J_O+J_I+J_C} \left[\mathbb{I}\{j \in \text{treatment group}\} \times t\right]^2} \\
&= \frac{\sum_{t=1}^{t^*-1} \sum_{j=1}^{J_O+J_I} t \left[\gamma_1 \cdot \mathbb{I}\{j \in \text{outlier}\} \times t + \epsilon_{jt}\right]}{(J_O + J_I) \sum_{t=1}^{t^*-1} t^2} \\
&= \gamma_1 \frac{\sum_{t=1}^{t^*-1} \sum_{j=1}^{J_O} t^2}{(J_O + J_I) \sum_{t=1}^{t^*-1} t^2} + \frac{\sum_{t=1}^{t^*-1} t \left(\frac{1}{J_O+J_I} \sum_{j=1}^{J_O+J_I} \epsilon_{jt}\right)}{\sum_{t=1}^{t^*-1} t^2} \\
&= \gamma_1 \frac{J_O}{J_O + J_I} + o_p(1).
\end{aligned} \tag{A.6}$$

If $\frac{J_I}{J_O} \to \infty$, then $\hat{\gamma} \to^p 0$ hence our estimate will converge to the true value and the test statistic will be properly centered around $\gamma = 0$. However, the condition $\frac{J_I}{J_O} \to \infty$ does not guarantee that we will obtain an unbiased estimate of the average treatment effect. To be precise, even if $\frac{J_I}{J_O} \to \infty$, the estimate of the average treatment effect will still be biased unless J_I/J_O grows faster than T. This implies that passing the pre-trends test is a necessary but not a sufficient condition for consistency of the treatment effect estimate when faced with outlier observations. Not only do we need the number of non-outlier units (J_I) to grow faster than the number of outlier units (J_O), we also need to have the former grow faster than the *total* number of outlier *observations* ($T \times J_O$).

A.2 Robust Regression Results Accounting for the Outlier Trend

Table A.4 provides additional robust regression results related to the discussion of Table 2 in the main text. The table shows various robust estimators, analogous to Table 1 but now accounting for the differential pre-trend for the ComEd observations—i.e., corresponding to Column (2) of Table 2. The regressions are similar to Equation (3), but now we also long-difference the pre- and post-averages for $\mathbb{1}(ComEd)_j \cdot \max[time_t - 35, 1]$. Once outliers are controlled for in Column (2), the robust estimators in Columns (3)-(6) become much more similar to the least-squares estimate. The ComEd plants are no longer outliers. The magnitude in Column (1), -12.5%, drops to -5.1% in Column (2) and the robust estimates range from -1.0% to -4.4%, corroborating the finding in Table 1 that the treatment effect is small and not statistically significantly different from zero.

Table A.4: Robust DiD Estimates of Log(Price) Accounting for the Outlier Trend

	(1)	(2)	(3)	(4)	(5)	(6)
	Least squares		Robust estimation			
			Q50	Bi.	S	MM
𝟙(Divest)	-0.125	-0.051	-0.010	-0.044	-0.027	-0.043
	(0.039)	(0.030)	(0.041)	(0.031)	(0.035)	(0.031)
𝟙(ComEd)		-0.007	-0.008	-0.007	-0.007	-0.007
·max[Time - 35,1]		(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
Drop ComEd	No	No	No	No	No	No
Synthetic-Level Clusters	174	174	174	174	174	174
Divested Plants	87	87	87	87	87	87
Control Plants	87	87	87	87	87	87
Breakdown Point			50		25	25
Gaussian Efficiency				95	76	95
R^2	0.055	0.493				
Observations	174	174	174	174	174	174

Notes : 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. *Q*50 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 for least-squares regressions, and at the synthetic control level for robust regressions.

A.3 Regression Results Using a Different Matching Criterion

This appendix shows that the main results of the paper do not hinge on the choice of matching on incentive regulation prior to divestiture. Tables A.5-A.6 repeat Tables 1-2 in the main text, but with "additional" matching that includes pre-divestiture regulatory status of each plant to distance and coal type in 1997. Figure A.5 shows the distribution of the plant-level robust statistics (analogous to Figure 1 in the main text). The estimates and figure are very similar to the results shown in the main text. Note that, with additional matching, one control plant is also flagged as an outlier. In Table A.6, Columns (2), (4) and (6), we also drop this outlier in the control group, which hardly affects the estimates but makes them ever so slightly smaller.

Table A.5: Plant Level Robust Analysis with Additional Matching

(a) DiD Estimates of Log(Price)

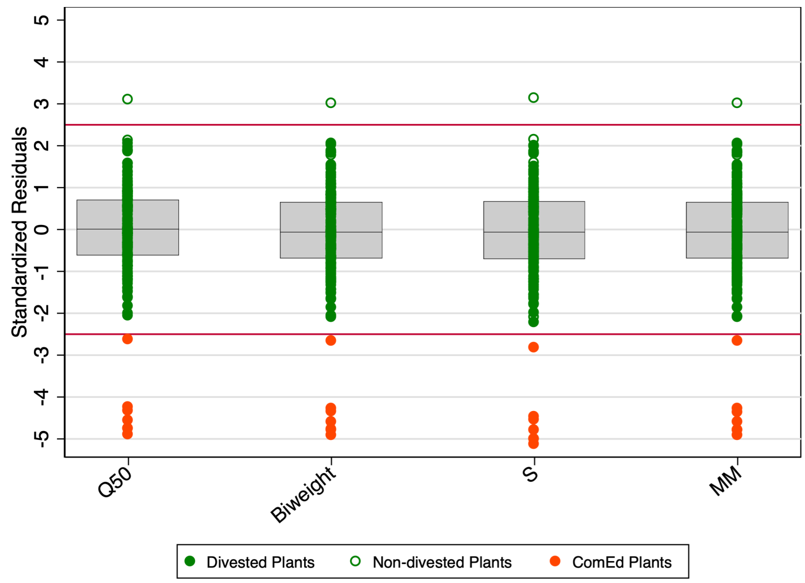
	(1)	(2)	(3)	(4)	(5)	(6)
	Least squares		Robust estimation			
			Q50	Bi.	S	MM
$\mathbb{1}$(Divest)	-0.117	-0.107	-0.019	-0.025	-0.004	-0.025
	(0.048)	(0.040)	(0.046)	(0.032)	(0.037)	(0.033)
	[0.093]					
Data	Monthly	Long-diff	Long-diff	Long-diff	Long-diff	Long-diff
Avg. # of Matched Neighbors	4.2	4.0	4.0	4.0	4.0	4.0
Year-Month FE	Yes	No	No	No	No	No
Plant FE	Yes	No	No	No	No	No
Plant-Level Clusters	168					
Utility-Level Clusters	78					
Synthetic-Level Clusters		162	162	162	162	162
Divested Plants	81	81	81	81	81	81
Control Plants	87	81	81	81	81	81
Breakdown Point			50		25	25
Gaussian Efficiency				95	76	95
Scale Estimate (RMSE)	0.191	0.258	0.205	0.207	0.207	0.207
R^2	0.638	0.042				
Observations	30,870	162	162	162	162	162

(b) Robust Statistics

	Divested	Not Divested	Total
$\sum_j \mathbb{1}(\lvert z_j\rvert > 2.5)$	7	1	8
$\sum_j z_j \cdot \mathbb{1}(\lvert z_j\rvert > 2.5)$	-30.4	3.0	-27.4

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, coal type in 1997 and pre-divestiture regulatory status of each plant. 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).

Figure A.5: Distribution of Plant Level Robust Statistics with Additional Matching



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

Table A.6: DiD Estimates of Log(Price) Accounting for the Outlier Trend with Additional Matching

	(1)	(2)	(3)	(4)	(5)	(6)
	No trend, drop ComEd		ComEd-specific trend		De-trended with ComEd projected post-trend	
𝟙(Divest)	-0.055	-0.055	-0.041	-0.039	-0.056	-0.054
	[0.055]	[0.055]	[0.054]	[0.054]	[0.056]	[0.056]
𝟙(ComEd) ·max[Time - 35,1]			-0.006	-0.006		
			[0.000]	[0.000]		
Drop ComEd	Yes	Yes	No	No	No	No
Drop Remaining Outlier	No	Yes	No	Yes	No	Yes
Avg. # of Matched Neighbors	4.2	4.2	4.2	4.2	4.2	4.2
Additional Matching	Yes	Yes	Yes	Yes	Yes	Yes
Year-Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes	Yes	Yes
Utility-Level Clusters	76	76	78	78	78	78
Divested Plants	74	74	81	80	81	80
Control Plants	81	81	87	87	87	87
R^2	0.711	0.711	0.716	0.711	0.699	0.694
Observations	28,288	28,288	30,870	30,740	30,870	30,740

Notes: Dependent variable is the logarithm of the coal price. Matching is based on distance, coal type in 1997 and pre-divestiture regulatory status of each plant. 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 Columns (5) and (6), dependent variable is de-trended using ComEd's estimated pre-trend and projected onto the post-period, as described in the main text.

A.4 Time Trends for Other Utilities

We now show a series of robustness checks in which we investigate the impact of introducing time trends for other utilities instead of only for ComEd, as well as the impact of leaving out other utilities.

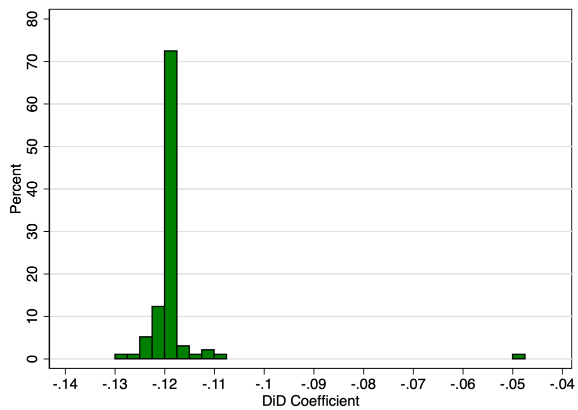
First, we include trends for other utilities one at a time. We estimate a DiD coefficient 98 times (i.e., for the 35 divested and 63 non-divested utilities), each time adding a linear trend for a single utility.[3] Panel (a) of Figure A.6 plots the distribution of the DiD coefficients. We find that the DiD coefficients are centered around -0.12, except when the trend is added for ComEd. The coefficient becomes -0.05 only when a trend is included for ComEd. As an additional check, in Panels (b) and (c), we estimate leave-one-out estimators at the utility and at the plant level. The utility-level leave-one-out estimator shows a similar distribution as in Panel (a), implying that ComEd is the only outlier with enough impact to shift the DiD coefficient substantially away from -0.12. The plant-level leave-one-out estimator reported in Panel (c) shows that unless a set of correlated outliers are dropped altogether (e.g., all plants belonging to a single utility), the estimated coefficient remains close to -0.12.

We then run a single regression with time trends for all utilities (including ComEd) and report the results in Table A.7. In Column 1, we report the baseline effect of -12%. In Column 2, we report the specification with the piecewise linear trend for ComEd only (this is Table 2, Column 2). In Column 3, we apply a piecewise linear trend for ComEd and a linear trend for all others, because other utilities do not necessarily have a contract renegotiation similar to ComEd at the end of 1992. Finally, in Column 4, we include linear trends for all utilities, including ComEd. The specifications in Columns 3 and 4 show that the estimated coefficients are similar to the coefficient in Column 2, where a piecewise trend is included only for ComEd. Had other utilties undergone a contract renegotiation like ComEd in the pre-period, the utility-specific trends would have corrected for such adjustments and shifted the DiD coefficient away from -0.050.

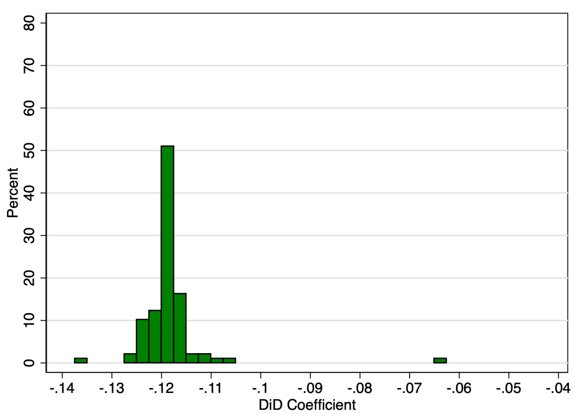
Taken together, these results suggest that ComEd is truly driving the results; other utilities do not have a large impact on the estimate.

[3] We also estimated a version including the piecewise linear trend that we use for ComEd in Column 2 of Table 2 for each utility. Results are unchanged and available from the authors upon request.

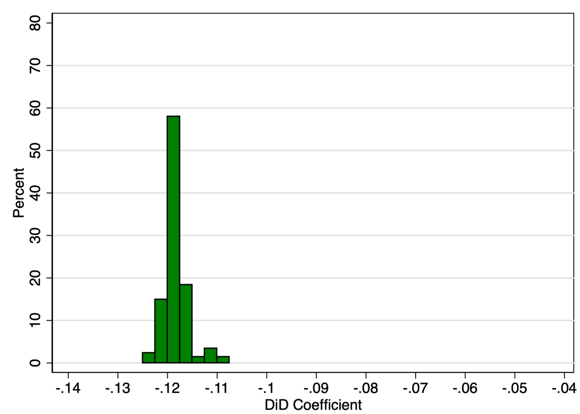
Figure A.6: Distribution of DiD Coefficients with Time Trends for, or Dropping, Other Utilities



(a) Utility-Specific Linear Trend



(b) Utility-level Leave-One-Out Estimator



(c) Plant-level Leave-One-Out Estimator

Notes: Figures show the distribution of DiD coefficients from 98 separate regressions in which a time trend is added for each utility one at a time (Panel (a)) or leaving out a utility one at a time (Panel (b)). In Panel (c), each plant is left out one at a time.

Table A.7: DiD Estimates of Log(Price) Accounting for Utility Trends

	(1)	(2)	(3)	(4)
	No trend	Piecewise trend for ComEd	Piecewise trend for ComEd; linear trend for others	Linear trend for all utilities
1(Divest)	-0.118	-0.050	-0.042	-0.049
	[0.082]	[0.045]	[0.035]	[0.039]
1(ComEd)		-0.006	-0.006	
·max[Time - 35,1]		[0.000]	[0.000]	
1(ComEd)·Time				-0.005
				[0.000]
Avg. # of Matched Neighbors	6.5	6.5	6.5	6.5
Year-Month FE	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes
Utility-Level Clusters	98	98	98	98
Divested Plants	87	87	87	87
Control Plants	119	119	119	119
R^2	0.720	0.790	0.795	0.794
Observations	38,093	38,093	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.

A.5 External Validity: The Choice of Coal Type and the Acid Rain Program

In this appendix, we provide a decomposition that is relevant in light of concerns related to the external validity of the full average treatment effect, as discussed in Section 3.2 of the main text. Divestiture was followed by a strong upward secular trend in the relative price of bituminous versus sub-bituminous coal after 2001, a period which also coincided with the implementation of the Acid Rain Program (ARP); a sulfur dioxide emissions permit trading program. Like divestiture, the ARP affects the type of coal that plants buy: plants are incentivized to choose sub-bituminous coal with lower sulfur content. Prior research has established that regulated monopolies in the control group had lower incentives to switch to low sulfur coal in response to the ARP, and were more likely to invest in "scrubber" technologies (Fowlie, 2010). In our sample divested plants are 10% more likely to use sub-bituminous coal.

Figure A.7 plots the relative price of the two types of coal over our sample period, along with the number of divested plants. Deregulation coincided with a pronounced aggregate shock in favor of divested plants: sub-bituminous coal became cheaper in relative terms from 1998 onwards but especially in 2001, as mining costs of bituminous coal increased due to the depletion of coal mines in the Eastern United States. Deregulation largely precedes that price shock, and there is no obvious reason why deregulation caused this trend. In addition, since divested firms also had reasons to switch to cleaner sub-bituminous coal for ARP compliance reasons, the ARP reinforced the effect of divestiture on coal type choice. Therefore, the concern is that the treatment effect of divestiture estimated in Table 2 in the main text is inflated—absent the large relative price shock and the additional incentives for coal switching introduced by the ARP, the treatment effect of divestiture on fuel cost would have been smaller. Moreover, in other contexts (e.g., other countries or periods), there is no reason to believe that divestiture would typically coincide with sulfur policies and price shocks; as such—other things equal—the United States' experience likely represents an upper bound on the cost savings from divestiture.

Figure A.7: Trends in Coal Prices by Type

Notes: Difference between the delivered cost for (high heat content) bituminous and (low heat content) sub-bituminous coal in $/MMBtu is plotted on the left vertical axis. The cumulative number of divested plants is shown on the right vertical axis. The average coal price between 1990 and 1997 is $1.54 for divested plants and $1.37 for non-divested plants.

We now proceed by estimating a lower bound on the treatment effect that we will argue is more likely to reflect the effect of divestiture in the absence of the simultaneous introduction of the ARP. Table A.8 isolates a more externally valid treatment effect. We re-estimate the average treatment effect from Table 2 in the main text, after controlling for the choice of coal type (i.e. "sub-bit" or "bit"), using the following regression:

$$\log(price)_{jt} = \alpha \mathbb{1}(divest)_{jt} + \beta_0 \cdot \mathbb{1}(sub - bit)_{jt} + \beta_1 \mathbb{1}(ComEd)_j \cdot \max[time_t - 35, 1] + \gamma_j + \delta_t + \epsilon_{jt} \quad \text{(A.7)}$$

where $\mathbb{1}(sub - bit)_{jt} = 1$ if plant j's majority coal type is sub-bituminous at time t. This decomposes the full causal treatment effect of divestiture into two channels. First, divested firms can negotiate better prices for a particular type of coal by renegotiating or searching for cheaper contracts and suppliers (the "direct effect" of divestiture). Second, divested firms can switch to a cheaper type of coal (the "indirect effect"). Table 2 presents the combination of these two effects (the full treatment effect); Table A.8 isolates the direct effect of divestiture that shuts down any cost savings from coal type switching—and thus represent a lower bound on the full treatment effect.[4] This comparison yields that there is a gap between the full vs. the direct effect of divestiture (-6.2% vs. -3.3% in our preferred specification), suggesting that divested plants achieved roughly equal cost savings through better negotiation for the same type of coal and through coal type switching. Neither the full nor the direct treatment effect of divestiture is statistically different from zero.

An alternative approach controls for heat content instead of for coal type. Because coal type is highly

[4]As coal type is a choice, we control for it to isolate channels rather than to obtain a consistent estimate of the full average treatment effect which would be subject to concerns about a "bad" endogenous control. However, we present additional evidence below to suggest that in our application the coefficient on $\mathbb{1}$(Divest) in Table A.8 is likely a more externally valid estimate of the full treatment effect of divestiture than the estimates in Table 2.

Table A.8: DiD Estimates of Log(Price) Controlling for Coal Type and Heat Content

	(1)	(2)	(3)
	No trend, drop ComEd	ComEd-specific trend	De-trended with ComEd projected post-trend
Panel A: Coal type controls			
𝟙(Divest)	-0.034	-0.023	-0.043
	[0.039]	[0.039]	[0.044]
𝟙(Sub-bit)	-0.208	-0.218	-0.205
	[0.041]	[0.038]	[0.039]
𝟙(ComEd)·max[Time - 35,1]		-0.007	
		[0.000]	
Panel B: Heat content controls			
𝟙(Divest)	-0.051	-0.037	-0.052
	[0.043]	[0.043]	[0.046]
Log(Heat Content)	0.263	0.322	0.313
	[0.168]	[0.159]	[0.155]
𝟙(ComEd)·max[Time - 35,1]		-0.006	
		[0.000]	
Drop ComEd	Yes	No	No
Avg. # of Matched Neighbors	6.3	6.5	6.5
Year-Month FE	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes
Utility-Level Clusters	93	98	98
Divested Plants	80	87	87
Control Plants	104	119	119
R^2 (Panel A)	0.788	0.790	0.779
R^2 (Panel B)	0.779	0.780	0.770
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, as described in the main text.

correlated with heat content, controlling for the latter is, in essence, a way to control for coal type in a continuous manner. Moreover, a continuous measure of coal type reflects the fact that plants receive deliveries of both types of coal and may even blend different coal types (Ellerman et al., 2000). We thus also estimate a version of Equation (A.7) using the logarithm of the heat content instead of a coal type indicator:

$$\log(price)_{jt} = \alpha \mathbb{1}(divest)_{jt} + \beta_0 \cdot \log(heat\ content)_{jt} + \beta_1 \mathbb{1}(ComEd)_j \cdot \max[time_t - 35, 1] + \gamma_j + \delta_t + \epsilon_{jt} \tag{A.8}$$

where $\log(heat\ content)_{jt}$ denotes the logarithm of the heat content of the coal used by plant j at time t (in MMBtu/ton). Panel B in Table A.8 shows the results. The gap between the full vs. the direct effect of divestiture (-6.2% vs. -5.1% in our preferred specification) is now smaller than in Panel A, but we still find that divested plants achieved some cost savings by contracting coal with a different heat content.

The direct treatment effects reported in Table A.8 represent a lower bound on the full treatment effect that we believe is more likely to reflect the effect of divestiture in the absence of the simultaneous introduction of the ARP. Fuel savings achieved for coal of the same type (the direct effect) cannot be attributed to the ARP, but cost savings through coal type switching could be.

Table A.9: DiD Estimates of the Effect of Divestiture on Heat Content and Coal Type

	(1)	(2)	(3)	(4)
	LHS: 𝟙(Sub-bit)		LHS: log(Heat content)	
𝟙(Divest)	0.107	0.132	-0.038	-0.059
	[0.056]	[0.067]	[0.016]	[0.018]
𝟙(Divest)		-0.095		0.075
·𝟙(Scrubber Before 97)		[0.070]		[0.021]
Avg. # of Matched Neighbors	6.5	6.5	6.5	6.5
Year-Month FE	Yes	Yes	Yes	Yes
Plant FE	Yes	Yes	Yes	Yes
p-value Sum of Coef.		0.476		0.330
Utility-Level Clusters	98	98	98	98
Divested Plants	87	87	87	87
Control Plants	119	119	119	119
R^2	0.790	0.792	0.925	0.928
Observations	38,093	38,093	38,093	38,093

Notes: Dependent variable the logarithm of the heat content in Columns (1) and (2) or a coal type indicator in Columns (3) and (4). Matching is based on distance and coal type in 1997. Maximum nearest neighbors is 10. Standard errors are clustered at the utility level are in square brackets.

We now present additional analysis that suggests that, absent the ARP, divestiture would have hardly affected the choice of coal type or heat content. To see this, we further investigate what drives the causal link between divestiture, environmental regulation, and the choice of coal type in Table A.9. We regress a plant's heat content or its choice of coal type (an indicator for sub-bituminous coal) on the DiD indicator $\mathbb{1}(divest)_{jt}$:

$$y_{jt} = \alpha_0 \mathbb{1}(divest)_{jt} + \alpha_1 \mathbb{1}(divest)_{jt} \cdot \mathbb{1}(scrubber\ before\ 1997)_{jt} + \gamma_j + \delta_t + \epsilon_{jt} \tag{A.9}$$

where y_{jt} is either $\log(heat\ content)_{jt}$ or $\mathbb{1}(sub-bit)_{jt}$. Column (1) excludes the interaction term and shows that divested plants are indeed more likely to burn low-sulfur sub-bituminous coal upon divestiture, which coincided with both the environmental regulation and relatively cheaper low-sulfur coal. Column (3) shows that the heat content of the coal used by divested plants went down by 3.7%.

Columns (2) and (4) present suggestive evidence that a divested plant's coal type or heat content is likely to be attributable to the environmental regulation, not divestiture. To establish to what extent divestiture would have affected coal type absent the Acid Rain Program, we investigate if plants that had a scrubber installed prior to 1997 were less likely to burn sub-bituminous coal than plants that did not. A scrubber is a device that can remove sulfur dioxide from industrial exhaust streams such as power plant emissions. It typically removes over 95% of SO_2 (Ellerman et al., 2000). Hence, plants with a scrubber have little incentive to choose lower heat-content sub-bituminous coal for environmental reasons (burning low-sulfur coal will not lead to much further SO_2 abatement once a scrubber is installed). If they choose that type of coal, the only objective must be to reduce fuel cost. In contrast, plants without a scrubber face both environmental and cost savings incentives to switch to a different type of coal.

Column (2) estimates Equation (A.9) and shows that divested plants with scrubbers installed by 1997 (i.e., plants that have few incentives to burn low-sulfur coal for ARP compliance but that may still have incentives to switch to sub-bituminous coal for cost-saving purposes) do *not* show a strong reduction in heat content, whereas divested plants without scrubbers do. Divested plants without scrubbers increase their use of sub-bituminous coal by 13.4 percentage points and show a heat content reduction of 5.9% post-divestiture. In contrast, divested plants with scrubbers increase their use of sub-bituminous coal only by 3.7 percentage points and *increase* their heat content by 1.6%, though neither of these effects is statistically significant.

The results in Table A.9 strongly suggest that, absent the ARP, divestiture would have impacted coal type and heat content modestly or not at all—i.e., the indirect effect of divestiture would have been near-zero and the smaller direct effect from negotiating better terms on contracts for the same coal type is more appealing from an external validity perspective.