Online Appendix: Project Networks and Reallocation Externalities

Vibhuti Dhingra
Schulich School of Business, York University, Toronto, ON M3J 1P3, Canada vibhutid@schulich.yorku.ca

Harish Krishnan
University of British Columbia, Vancouver, BC V6T 1Z4, Canada Harish.Krishnan@sauder.ubc.ca

Juan Camilo Serpa
Desautels Faculty of Management, McGill University, Montreal, QC H3A 1G5, Canada juan.serpa@mcgill.ca

Appendix A: Robustness analyses

A.1. Alternative delay metrics

Our models thus far use two delay metrics: $\textit{Delay Days}_{p,t}$ (a continuous variable that measures the length of the delay), and $\textit{Delay Probability}_{p,t}$ (where the dependent variable is an indicator that equals one if there was a reported delay, and is zero otherwise). In this section, we rerun our analysis with two alternative delay measures: $\text{Relative Delay}_{p,t} = 100 \times \frac{\text{Delay Days}_{p,t}}{\text{Initial Duration}_p}$ (following Calvo et al. 2019) and $\text{Percentage Delay}_{p,t} = 100 \times \frac{\text{Delay Days}_{p,t}}{\text{Deadline}_{p,t-1} - \text{Start Date}_p}$. The first measure benchmarks the reported delay days to the baseline project duration, whereas the second measure benchmarks the number of delay days as a function of the current projected execution time (i.e., the expected deadline minus the start date). Table A.1 shows that the results are still significant using these alternative metrics.

Table A.1 Alternative dependent variables

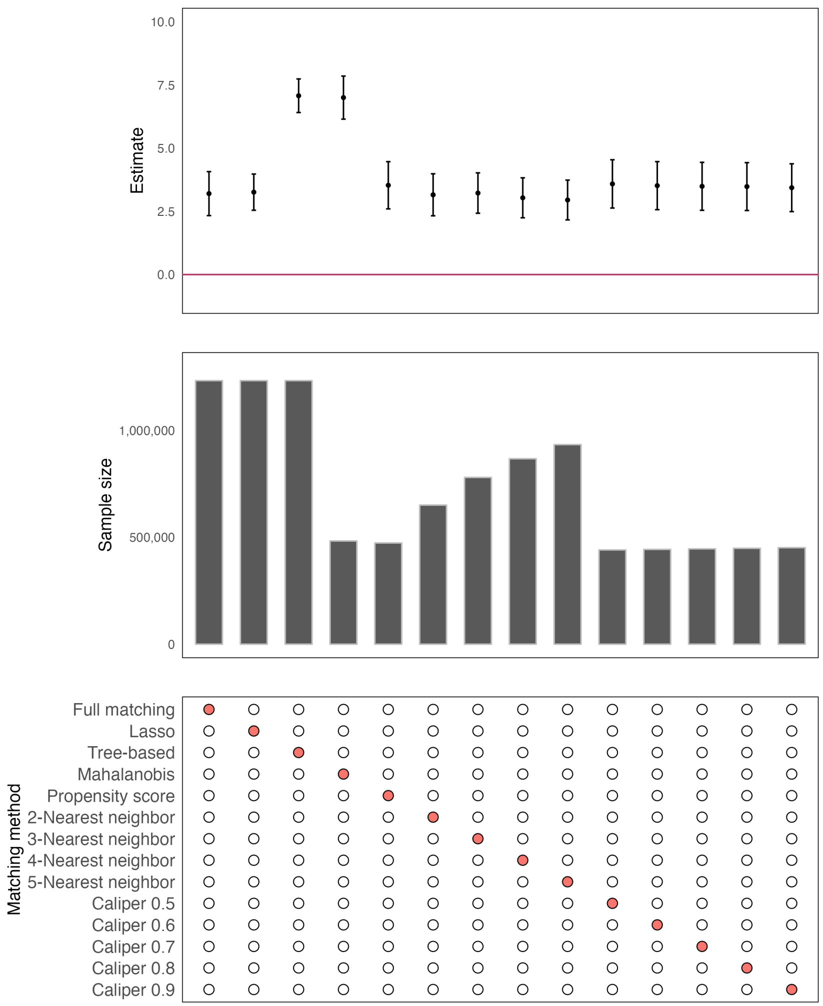
	Agency nodes						Contractor nodes					
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII
Dep. variable:												
Relative delay	0.78***	0.46**	0.47**	0.73***	0.85***	1.41***	1.36***	1.55***	1.27***	1.44***	1.01***	1.85***
	(0.16)	(0.20)	(0.23)	(0.22)	(0.23)	(0.29)	(0.25)	(0.27)	(0.31)	(0.31)	(0.32)	(0.45)
Percentage delay	0.84***	0.62***	0.59***	0.87***	0.98***	1.46***	1.07***	1.11***	0.88***	1.18***	0.82***	1.58***
	(0.13)	(0.16)	(0.18)	(0.18)	(0.18)	(0.25)	(0.19)	(0.21)	(0.24)	(0.24)	(0.27)	(0.39)
Agency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contractor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Matching specification:												
Number of bins	2	2	3	4	5	Sturges	2	2	3	4	5	Sturges
Location	No	State	State	State	State	State	No	State	State	State	State	State
Identifiable partitions	22,219	49,299	52,508	52,461	51,412	29,884	19,158	45,183	46,756	45,944	44,674	24,006
Observations	1,118,711	706,317	583,399	491,845	432,644	139,531	1,087,062	666,371	539,473	445,100	391,093	117,750
R^2	0.26875	0.28503	0.29281	0.30161	0.30566	0.35856	0.14499	0.16632	0.17381	0.18001	0.18277	0.20990

Note. This table presents the estimated coefficients for the spillover effect of a network disruption on delay relative to a project's initial duration, and on delay relative to a project's current duration. Columns I-VI show the effect for projects connected through the agency nodes, and Columns VII-XII show the effect for projects connected through the contractor nodes. Treated and control projects are always matched on two-digit task code, two-digit industry code, and four price categories, and numerical variables (number of bids, initial budget, initial duration, annual revenue, and number of employees) using different levels of coarsening. In Columns I and VII, numerical variables are coarsened into two bins. Columns II-VI and VIII-XII gradually increase the number of bins yielding finer partitions. All specifications include fixed effects for the agency, contractor, and project task. The standard errors (reported in parentheses) are robust and clustered at the project level. Significance levels: 10% (*), 5% (**), and 1% (***)

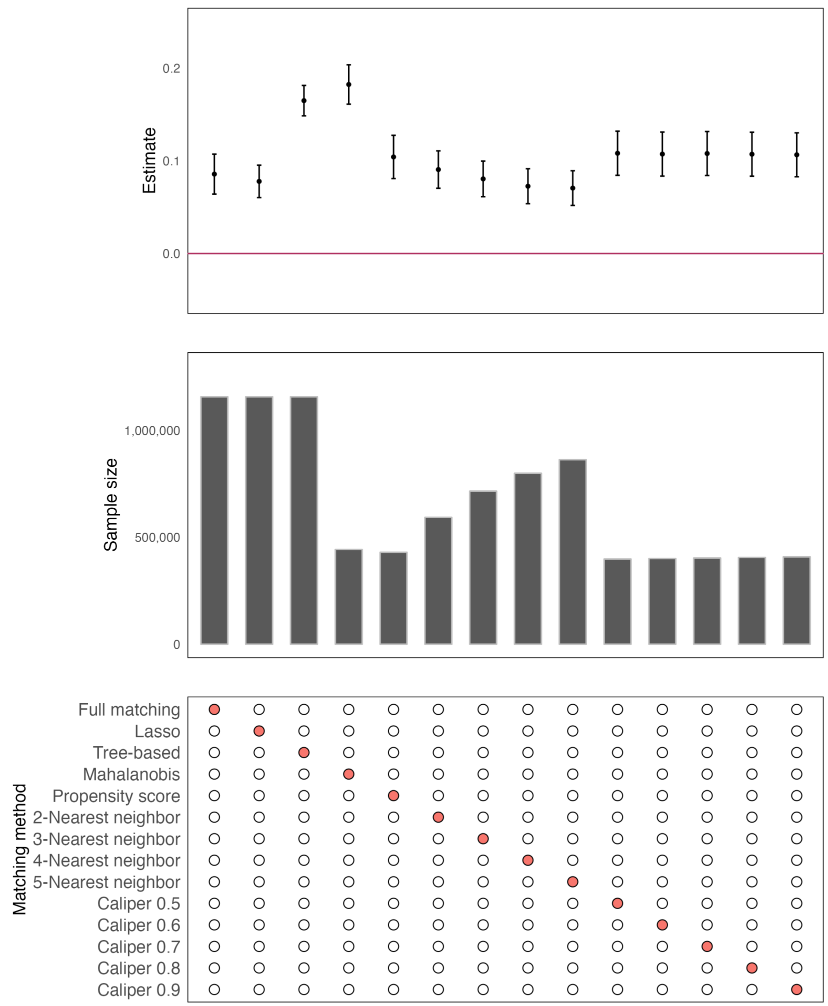
A.2. Alternative matching methodologies

In our main analysis, we employed Coarsened Exact Matching to partition projects. In this section, we supplement our analysis by changing the matching methodology to other popular methodologies. In particular, we re-estimated our main results by using propensity score matching; Lasso-based matching; Tree-based matching; Mahalanobis distance matching; n-caliper matching, with five different calipers ($n = 0.1, 0.2, 0.3, 0.4, 0.5$); and k-nearest-neighbor matching (k-NN), where we let $k = 2, 3, 4, 5$.

Figures A.1 and A.2 show that our results are robust across all methods, both in sign and significance. Note that the figures only report the results for the regression specification that includes all controls and fixed effects. However, the regression results were robust across all other specifications.

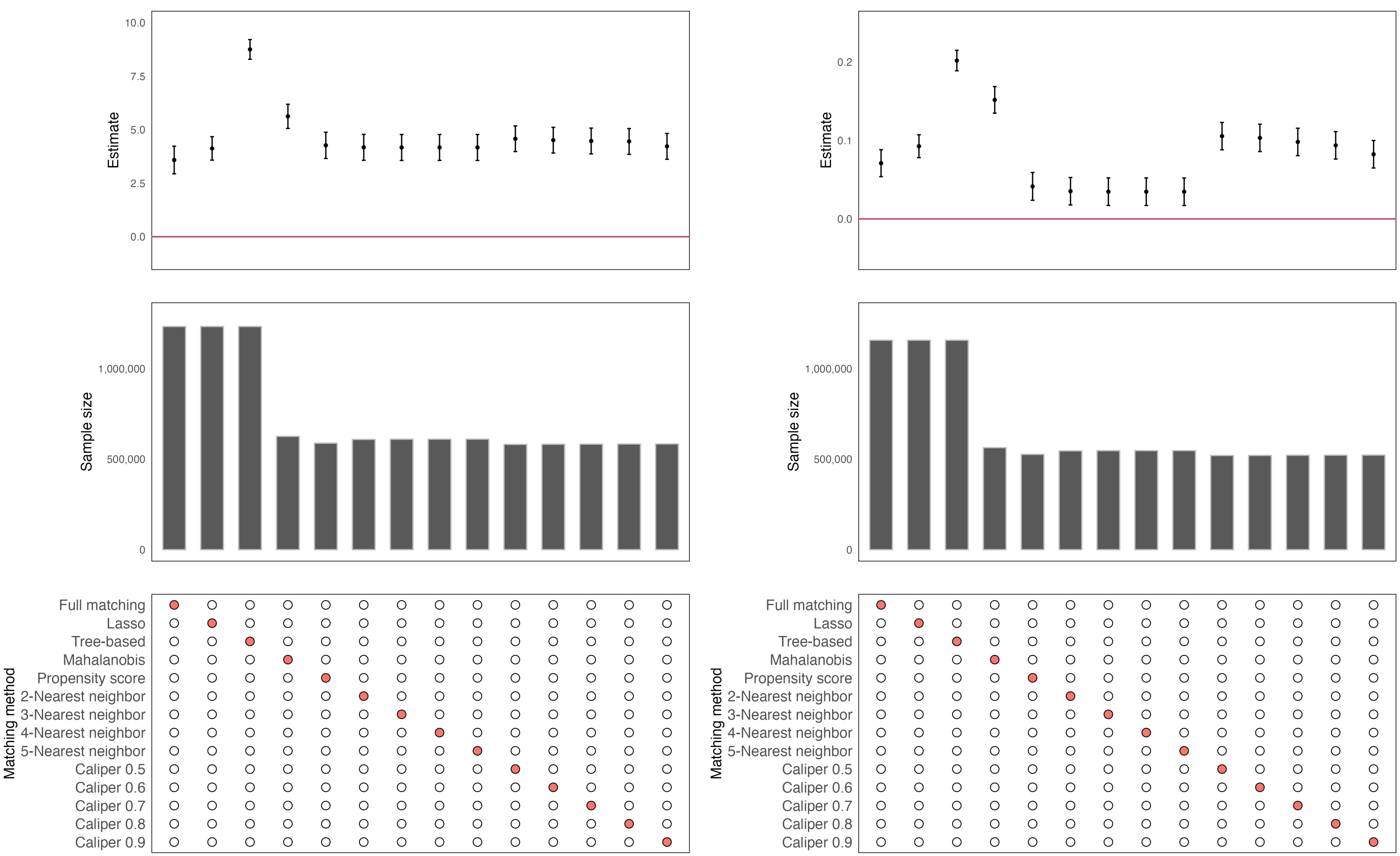


(a) Delay days



(b) Delay probability

Figure A.1 Contractor nodes – Spillover effect with alternative matching methodologies.
These figures show our main results for spillover effect along contractor nodes under different matching techniques. For instance, the first column in each figure shows the results from optimal full matching whereas the last column shows the results from propensity score matching with 0.9 caliper size. The bottom panel illustrates the matching method used (shown in red colored bullets). The middle panel displays the number of observations in each matched sample. The top panel displays the estimated treatment effect and its 95% confidence interval.

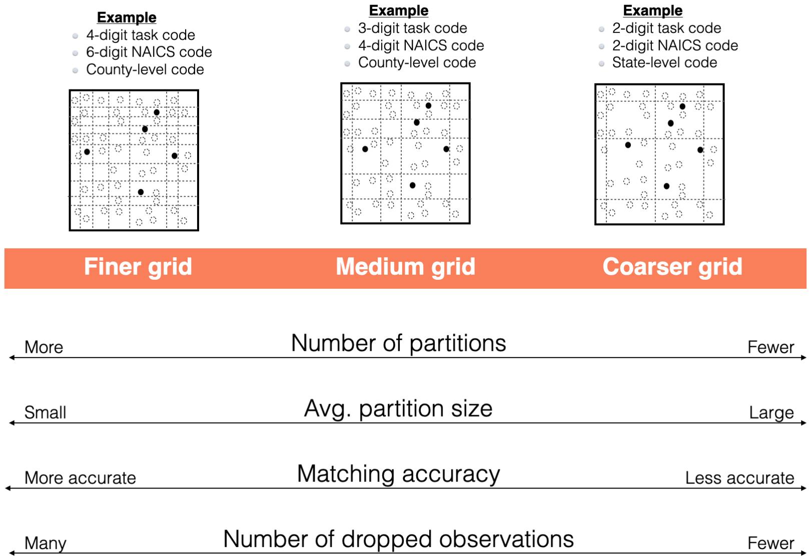


(a) Delay days (b) Delay probability

Figure A.2 Agency nodes – Spillover effect with alternative matching methodologies.

These figures show our main results for spillover effect along agency nodes under different matching techniques. For instance, the first column in each figure shows the results from optimal full matching whereas the last column shows the results from propensity score matching with 0.9 caliper size. The bottom panel illustrates the matching method used (shown in red colored bullets). The middle panel displays the number of observations in each matched sample. The top panel displays the estimated treatment effect and its 95% confidence interval.

Figure A.3 Partition granularity



A.3. Exact matching with coarser partitions

As mentioned in §4, our dataset allows us to match projects at different levels of granularity. We could, for example, partition the sample using the 4-digit or 5-digit NAICS code (instead of the 6-digit code), or the 2-digit or 3-digit task-level code (instead of the four-level digit). We could also partition at the county level, instead of requiring it to be at the state level.

Partitioning on higher level variables will result in broader (but fewer) partitions with matches that are slightly less precise. However, it will allow us to utilize most of the observations in our analysis, thereby yielding a more representative estimation (see Figure A.3 for an example). In Table A.2, we present regressions in which we progressively altered the granularity of the partition, transitioning from a very fine to a very coarse partition scheme. As can be seen, the results are largely unaffected by this variation, indicating that our findings are not solely determined by the choice of partition construction.

Table A.2 Exact matching with coarser partitions

	Agency nodes							Contractor nodes						
	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	XIII	XIV
Dep. variable:														
Delay probability	0.11***	0.11***	0.12***	0.11***	0.11***	0.09***	0.09***	0.11***	0.11***	0.10***	0.09***	0.09***	0.08***	0.05***
	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)
Delay days	4.02***	4.05***	4.50***	4.33***	4.20***	3.79***	3.36***	5.72***	5.56***	5.30***	5.09***	5.15***	4.20***	2.06***
	(0.86)	(0.83)	(0.74)	(0.72)	(0.65)	(0.58)	(0.47)	(0.97)	(0.94)	(0.87)	(0.85)	(0.80)	(0.72)	(0.64)
Partition granularity:														
NAICS code	5-digit	4-digit	3-digit	2-digit	2-digit	2-digit	2-digit	5-digit	4-digit	3-digit	2-digit	2-digit	2-digit	2-digit
Task code	4-digit	4-digit	4-digit	4-digit	3-digit	2-digit	2-digit	4-digit	4-digit	4-digit	4-digit	3-digit	2-digit	2-digit
Location	County	County	County	County	County	County	State	County	County	County	County	County	County	State
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Agency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contractor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Identifiable partitions	29,405	30,698	33,414	35,239	37,847	41,710	39,058	32,301	33,334	36,442	38,654	41,229	44,266	37,399
Observations	225,965	242,360	294,445	313,043	378,271	512,381	844,315	255,910	270,310	321,144	341,267	403,883	529,356	834,023
R^2/Pseudo R^2	0.31304	0.31281	0.30370	0.30254	0.30653	0.29567	0.29635	0.18021	0.17703	0.18239	0.17927	0.17129	0.17866	0.16563

Note. This table reports the estimated treatment effect of network disruptions on delay probability and delay days using coarser variables in matching. Columns I-VII report the results for agency nodes and Columns VIII-XIV report the results for contractor nodes. All regression specifications include fixed effects for the contractor, agency, task, and price. Standard errors (in parentheses) are robust and clustered by project. Significance levels: $1\%^{***}, 5\%^{**}, 10\%^{*}$.

Table A.3 Statistics for the unadjusted and adjusted distributions of the data

		Standardized bias		Variance ratio		Kolmogorov-Smirnov	
Covariate	Type	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
Annual Revenue – Contractor	Continuous	0.04	0.00	1.28	1.00	0.03	0.03
Number of Employees – Contractor	Continuous	0.04	0.00	1.25	1.00	0.04	0.03
Number of bids	Continuous	0.11	0.00	1.29	1.01	0.04	0.00
Project budget	Continuous	0.04	0.01	1.09	1.00	0.02	0.03
Project duration	Continuous	0.01	0.01	1.24	1.01	0.05	0.01
Competitively awarded contract	Categorical	0.01	0.01	·	·	0.01	0.01
Performance-based Incentives	Categorical	0.02	0.01	·	·	0.02	0.01
Cost contract	Categorical	0.01	0.00	·	·	0.01	0.00
Fixed price contract	Categorical	0.03	0.00	·	·	0.03	0.00
Labor hours contract	Categorical	0.01	0.00	·	·	0.01	0.00
Time and materials contract	Categorical	0.01	0.00	·	·	0.01	0.00

A.4. Balancing tests of the matched samples

In our main analysis, we matched treated and control observations using coarsened exact matching. In this section, we verify that, after matching, the treatment and control observations have similar distributional properties across the matching covariates. We verify distributional balance across three statistics of the treatment and control samples: (i) the means; (ii) the variances; and (iii) the cumulative distributions (Stuart et al. 2013):

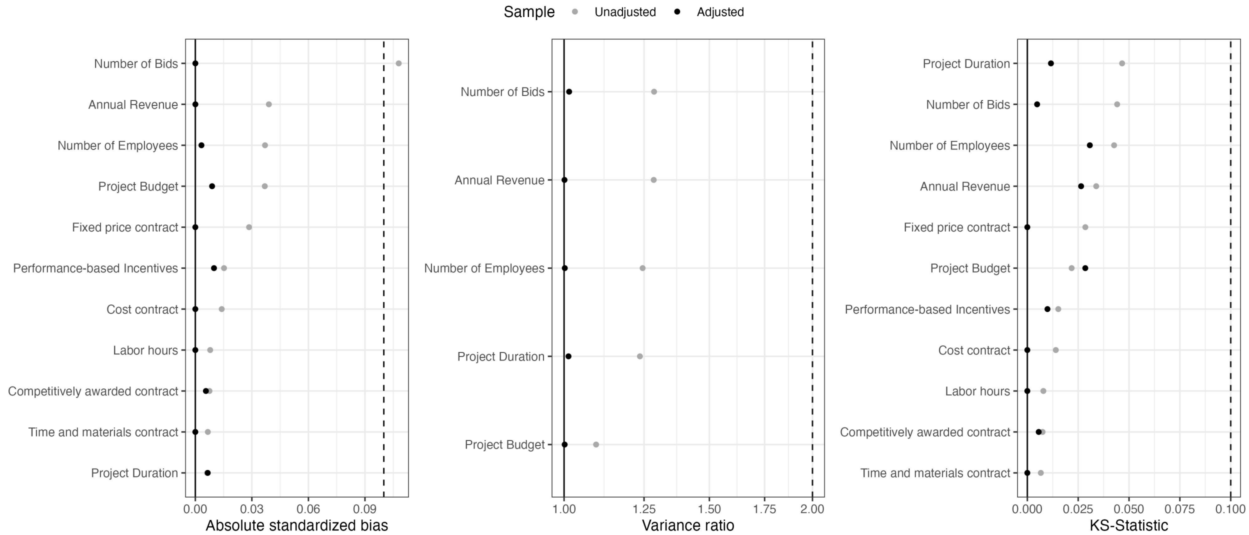
1. *Balance of means.* To measure the balance of the distributions' first moment, we measure the standardized bias for each variable. We calculate the standardized bias of covariate x by measuring the absolute difference of the means, $\left|\mu_x^{treatment} - \mu_x^{control}\right|$, and dividing this difference by the pooled standard deviation. As a rule of thumb, the adjusted standardized differences should be smaller than 0.1 (Stuart et al. 2013). Figure A.4 and Table A.3 illustrate the standardized bias for each covariate, before and after weighting the distributions. These two exhibits show that the distributional means are balanced.
2. *Balance of variances.* To explore the balance of the distributions' variances, we analyze the ratio of the treatment and control groups' variance. By convention, we place the largest variance in the numerator; a ratio of one means that the variances are perfectly balanced and, as a rule of thumb, a ratio below two is acceptable after adjusting the distributions. Table A.3 shows that the distributions' variances satisfy this requirement.
3. *Balance of cumulative distribution.* We explore the balance of the cumulative distribution functions via the Kolmogorov-Smirnov statistic, which measures the maximum distance between the support of these functions. This statistic ranges from zero (perfect balance) to one (full imbalance). By convention, a value below 0.05 is recommended after adjusting. Table A.3 shows that all our adjusted covariates meet this recommendation.

In summary, the balancing table results show that the treatment and control samples are balanced in their means, variances, and cumulative distributions. Figures A.5 and A.6 further illustrate that the density plots are balanced across several covariates for treated and control observations.

A.5. Placebo tests

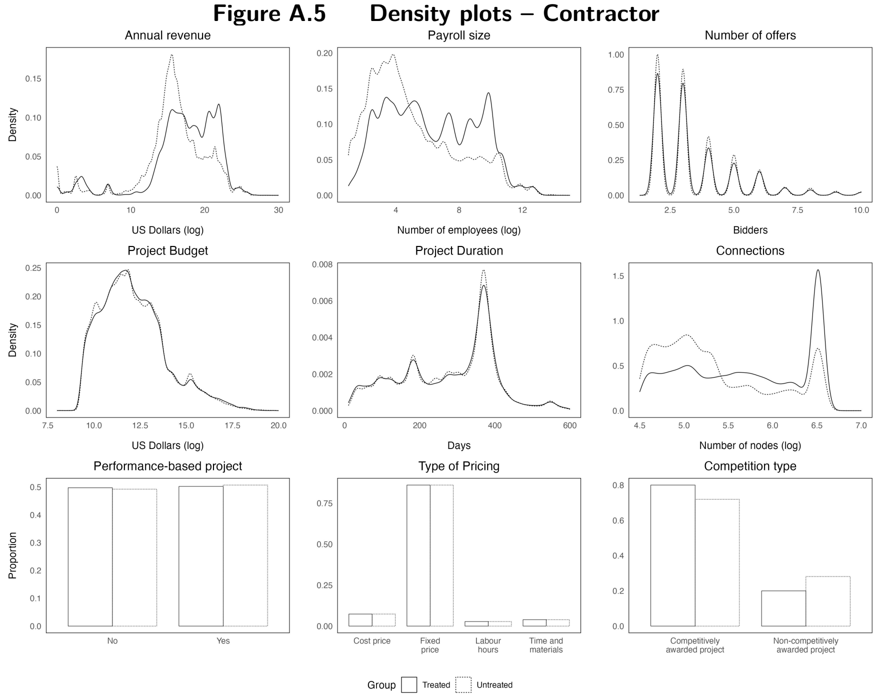
A.5.1. Inter-temporal placebo: Testing for parallel trends. Our estimates do not come from a prototypical diff-in-diff model, so testing for the parallel trends assumption is not straightforward. This is for

Figure A.4 Balancing test results



Note. This figure illustrates the distributional balance between treated and control observations after matching. The leftmost panel shows the balance of means, the middle panel shows the balance of variances, and the rightmost panel shows the balance of cumulative distribution functions.

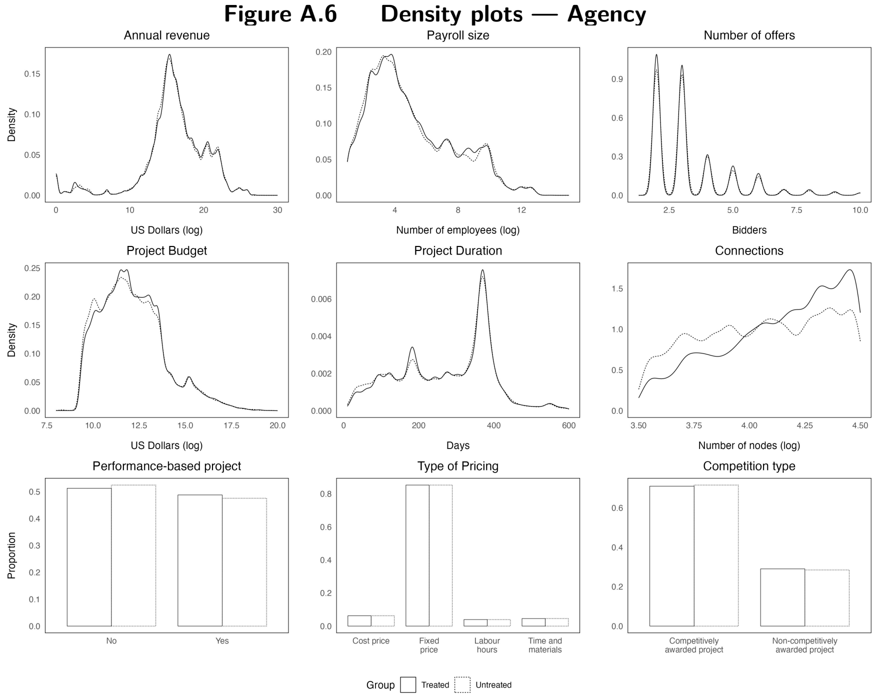
Figure A.5 Density plots – Contractor



Note. This figure illustrates the distribution plots of the treated and control observations for contractor nodes after matching.

three reasons: (i) observations in our sample can be 'treated' repeatedly (a given project can be disrupted at any point in the time series); (ii) the treatment is not coordinated along the cross-section (i.e., disruptions hit different projects in different periods); and (iii) the treatment effect dissipates across time (i.e., the impact of a disruption is not permanent).

Figure A.6 Density plots — Agency

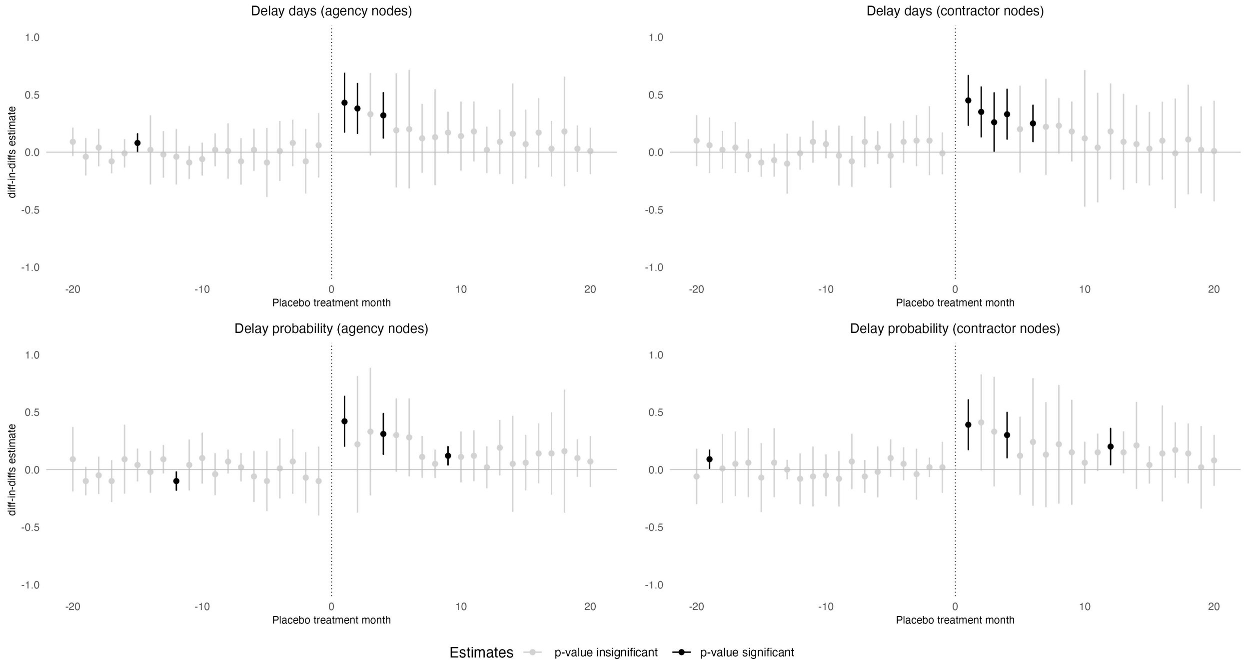


Note. This figure illustrates the distribution plots of the treated and control observations for agency nodes after matching.

These three peculiarities mean that any "off-the-shelf" parallel-trends test—like Autor (2003)'s placebo test or a lead-lag test—will not work in our setting. We can, however, still test the fundamental assumption of a diff-in-diff estimator—that absent a treatment, the outcome of the two groups would have been identical over time—by performing an adaptation of Monfared and Pavlov (2017)'s inter-temporal placebo test. In this test, we create 40 synthetic samples, each being identical to the true sample except that we pretend that the disruption occurred during a different period. Twenty of these datasets are called *lagged* synthetic samples, while the remaining twenty are called *lead* synthetic samples, where (i) the n^{th} lagged sample pretends that the treatment date occurred n periods *before* the true event date and (ii) the n^{th} lead sample pretends that the treatment date occurred n periods *after* the true event date. The idea is to shift the treatment date back and forth, one quarter-year period at a time (for a maximum of twenty periods each way), and then determine how this shift would affect the diff-in-diff coefficients. Figure A.7 plots the value of all forty placebo coefficients (i.e., Periods, -20,...,-2,-1,1, 2, ,..., 20) and their corresponding 95% confidence interval. For ease of interpretation, we normalize the value of the true coefficients to 1. According to Monfavred and Pavlov's test, if the coefficients are significant and the parallel-trends assumption is valid, then all lagged placebo regressions would be insignificant (i.e., there are no anticipatory trends). In contrast, the lead placebo regressions would be significant but the effect should dwindle as time passes. Figure A.7 confirms this pattern, by showing a lack of anticipatory effects and a lead treatment effect that lingers for two periods (or six months).

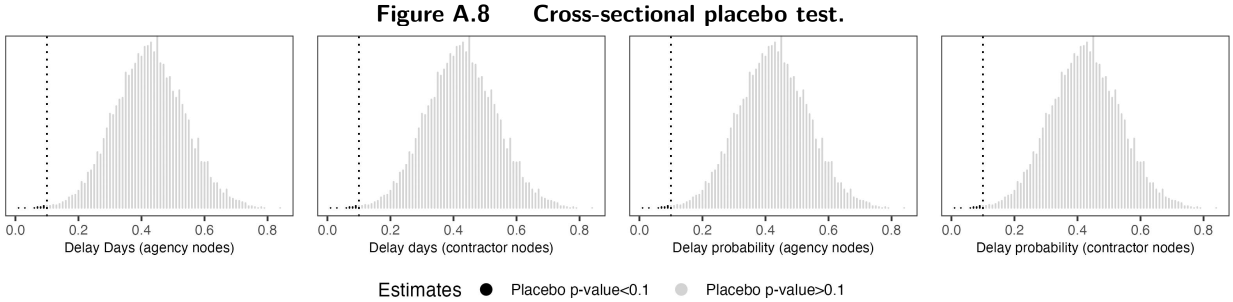
A.5.2. Cross-sectional placebo test. We also conduct a cross-sectional placebo test to determine if our results are artifacts of spurious correlations in the data. We create 10,000 samples by randomly assigning disruption occurrences across observations via Bernoulli trials. These 10,000 synthetic samples are

Figure A.7 Inter-temporal placebo test.



Note. For forty synthetic datasets, we re-estimate our models by artificially setting the disruption time period to be different from the "true" event date. Twenty of these samples set the "placebo" treatment date before the true event, and twenty set it to be after the true event. We plot the distribution of p-values for the diff-in-diff coefficient of the forty placebo estimates. When the p-value exceeds 0.1, it is shaded gray; if the draw of a placebo regression yields a p-value below 0.1, it is shaded black. All estimates are drawn from regression models including all fixed effects and control variables.

Figure A.8 Cross-sectional placebo test.



Note. For 10,000 synthetic datasets, we re-estimate our model specifications. We plot the distribution of p-values for the diff-in-diff coefficient of the 10,000 placebo estimates. When the p-value exceeds 0.1, it is shaded gray; if the draw of a placebo regression yields a p-value below 0.1, it is shaded black. Each of the four plots contains the distribution using a different dependent variable. All estimates are draw from regressions models including all fixed effects and control variables.

all identical to our true sample except that observations are randomly sorted into the "treated" and "control" groups. Figure A.8 presents the p-values of the 10,000 placebo estimates. If our true estimates were artifacts of spurious correlation, then a substantial number of the synthetic samples would have low p-values. But Figure A.8 shows that a vast majority of the placebo estimates are not even significant at the 10% level.

A.6. Alternative standard errors

In our main analysis, we accounted for serial correlation by clustering the standard errors at the project level consistent with the recommendation in Bertrand et al. (2004) and Cameron and Miller (2015). In this section, we assess the robustness of our findings to alternate levels of clustering. Table A.4 shows that our estimates are still positive and statistically significant when we cluster the standard errors at the contractor, agency, county, and task level, or use a combination of contractor and agency level clusters.

We also account for potential temporal correlation in the data, and consider alternative measures of standard errors. Table A.5 shows that our results are robust to these measures, including heteroscedasticity and autocorrelation consistent standard errors and heteroscedasticity-consistent standard errors—see Cameron and Miller (2015) for a detailed discussion of these issues.

Table A.4 Alternative cluster levels

	Agency nodes					Contractor nodes				
	I	II	III	IV	V	VI	VII	VIII	IX	X
Delay days	3.74***	3.74***	3.74***	3.74***	3.74***	4.32***	4.32***	4.32***	4.32***	4.32***
	(0.61)	(0.87)	(0.73)	(0.63)	(0.85)	(1.11)	(1.53)	(1.06)	(1.01)	(1.53)
Cluster level	Contractor	Agency	County	Task	Agency + Contractor	Contractor	Agency	County	Task	Agency + Contractor
Agency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contractor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	491,846	491,846	491,846	491,846	491,846	445,101	445,101	445,101	445,101	445,101
R^2	0.26	0.26	0.26	0.26	0.26	0.11	0.11	0.11	0.11	0.11

Note. This table reports the estimated treatment effect of network disruptions on delay days by clustering the standard errors (in parentheses) at different levels. Significance levels: $1\%^{***}, 5\%^{**}, 10\%^{*}$.

Table A.5 Alternative standard errors

	Agency nodes				Contractor nodes			
	I	II	III	IV	V	VI	VII	VIII
Delay days	3.74***	3.74***	3.74***	3.74***	4.32***	4.32***	4.32***	4.32***
	(0.57)	(0.86)	(0.53)	(0.51)	(0.81)	(1.30)	(0.78)	(0.75)
SE Type	Newey-West	Driscoll-Kraay	Degree of Freedom	Bell-Mcaffrey	Newey-West	Driscoll-Kraay	Degree of Freedom	Bell-Mcaffrey
Agency FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Contractor FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Task FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	491,846	491,846	491,846	491,846	445,101	445,101	445,101	445,101
R^2	0.26	0.26	0.26	0.26	0.11	0.11	0.11	0.11

Note. This table reports the estimated treatment effect of network disruptions on delay days by considering alternative measures of the standard errors (reported in parentheses). Significance levels: $1\%^{***}, 5\%^{**}, 10\%^{*}$.

Appendix B: Measuring the economic costs of delay spillovers

The costs of a delay to a project can manifest in many ways: there can be direct, indirect, and opportunity costs. For example, if a traffic jam delays our evening commute home, there are direct costs (cost of gas), indirect costs (additional wear and tear on car) and opportunity costs (less time at home to spend with family or to relax). In other words, delay costs ripple outward – starting with costs that are easy to attribute to the delay and to estimate, and moving to indirect and opportunity costs that are harder, if not impossible, to estimate or even to identify or attribute to the original delay.

This idea, that delay costs ripple outward in the manner described above, is recognized in studies that attempt to estimate the economic significance of delays. For example, the Texas Dept of Transportation (Beaty et al. 2016) developed a model to estimate the costs associated with delay of highway projects.

In our setting, in order to estimate the costs of delay spillovers, we first start by conceptualizing the different categories of delay costs. Keogh and Evans (1992) provide a simple but useful categorization. Project delay costs can be classified as leading to "private costs" and "social costs." Private costs refer to the costs of delays that are incurred by the organizations that are directly involved (e.g., the contractors, agencies, etc.); these costs can be direct (e.g., additional labor costs) or indirect (e.g., overheads incurred while the project is delayed). Social costs refer to the costs of delay that are externalized, i.e., borne by society at large. These costs can include the impact on citizens (e.g., commuters when a road construction project is delayed), etc. Beaty et al. (2016) point out, however, that in public projects, all costs are ultimately borne by the general public.

There are some examples of studies that attempt to estimate social costs of a delay. Using a case study of highway repair projects, Lewis and Bajari (2014) argue that "the daily social cost imposed by the construction would be 175,000 hours. Valuing time at $10 an hour, this implies a social cost of $1.75 million per day."

In our context, it might be impractical, if not impossible, to attempt to quantify the social costs of delays. This is because of the wide variety of projects and users that are part of our data set. Nonetheless, we acknowledge that these costs can be significant. Keogh and Evans (1992) also note that social costs are extremely hard to estimate and that estimates of delay costs are often confined to the estimation of private costs. We follow a similar approach here and restrict our estimation of costs to the direct and indirect cost borne by the contractors. This estimate, of course, would be a lower bound on the true delay costs.

Measuring direct and indirect costs of project delays

A straightforward approach to estimating the direct costs of a delay is to simply look at the cost overrun (i.e., the actual project cost – the initial estimated project cost). The benefit of this approach is its simplicity. However, cost overruns may incorrectly estimate the direct costs, e.g., if some direct costs cannot be billed to the project.

For projects where labor costs are the major component of direct costs of delay, estimating the additional labor cost due to project delays provides a conservative estimate of delay costs.

Indirect costs are even more challenging to account for and estimate. This is because indirect costs are often overheads that a contractor allocates to its projects. How would a delay affect the allocation and recovery of the additional overhead costs that a contractor may incur?

U.S. Federal Courts have accepted a formula called the Eichleay formula as a valid method to estimate overhead costs that cannot be directly attributed to a particular project but are incurred due to the extended duration of the project caused by the delay. Overhead costs typically include administration costs, utilities, depreciation, taxes, insurance, etc., which are not directly billable to a specific project but are necessary business costs. The Eichleay formula is calculated via a three-step procedure:

- Step 1: Total Project Billings ÷ Total Company Billings × Total Home-Office Overhead During Actual Contract Period = Overhead Allocated to Project
- Step 2: Overhead Allocated to Project ÷ Actual Days of Project Performance (including delay) = Rate of Overhead Allocated to Project Per Day
- Step 3: Rate of Overhead Allocated to Project Per Day × Number of Days Delayed = Amount of Overhead Allocated to Project due to delay.

Using data to estimate the economic costs of delay spillover

We estimate the direct and indirect costs of a delay using three methods commonly used by the project managers and the courts to assess the damages caused by a delay.

1. Eichleay formula: For the first estimate, we use the labor cost of delay to estimate the direct cost and we estimate the indirect cost using the Eichleay formula.

The labor cost of delay (i.e., direct costs of delay) was estimated by retrieving the average hourly wage in every industry from the Bureau of Labor Statistics, and the average overhead cost and company billings from filing records. We, then, matched this with project records to estimate the cost of labor in the project's industry. We computed the labor cost for a project delay as follows:

$$\text{Direct cost of labor} = 8 \times \text{Hourly wage per worker} \times \text{Days of delay} \times \text{Employees per project}$$

We obtain the overhead allocated to a project by re-writing the first two steps of Eichleay formula:

$$\text{Daily overhead allocated to project} = \frac{\text{Total project cost}}{\text{Actual project duration}} \times \frac{\text{Total overhead}}{\text{Total company billings}}$$

Once the overhead allocated to a project is defined, the indirect costs are calculated as follows.

$$\text{Indirect costs} = \text{Daily overhead allocated to project} \times \text{Delay days}$$

2. Modified Eichleay formula: For the second estimate, we again use the labor cost of delay to estimate the direct cost and we estimate the indirect cost using a Modified Eichleay formula.

The modified Eichleay formula uses a slightly different formula for the estimating the overhead allocated to a project. The formula is as follows:

$$\text{Daily overhead allocated to project} = \frac{\text{Total project cost}}{\text{Initial project duration}} \times \frac{\text{Total overhead}}{\text{Total company billings}}$$

Note that the denominator in the modified Eichleay formula is the initial project duration and not the actual project duration. Again, the indirect costs are calculated by multiplying the daily overhead allocated to a project with delay days.

3. CD3 formula: A third approach we use to estimate delay costs is called the "cost of delay divided by duration" or "CD3". This is a commonly used approach by contractors, and it does not require much data. The CD3 approaches estimates the sum of direct and indirect costs as follows.

$$\text{Direct Cost} + \text{Indirect cost} = \frac{\text{Total project cost}}{\text{Initial project duration}} \times \text{Delay days}$$

Once we have been able to estimate the economic costs of a project delay, we can quantify the economic costs of a delay spillover. In order to do this, we first measure the size of spillovers per delay day. We then multiply the per-day total cost by this magnitude. Put differently:

$$\text{Economic costs of delay spillover} = \text{Cost of delay per day} \times \text{Spillover magnitude}$$

We adapt the above methods using our data. In particular, to estimate the direct cost of a delay, we retrieved the average hourly wage in every industry from the Bureau of Labor Statistics, and proxy for the average overhead cost and company billings using filing records from Compustat. We, then, matched this with project records to estimate the cost of labor in the project's industry. Note that this approach can only provide a lower bound on the direct cost of delays as it does not account for some inputs. However, our intent here is to illustrate a methodology which can be used to estimate delay costs, and a lower bound will at least provide some idea about the economic impact of a delay.

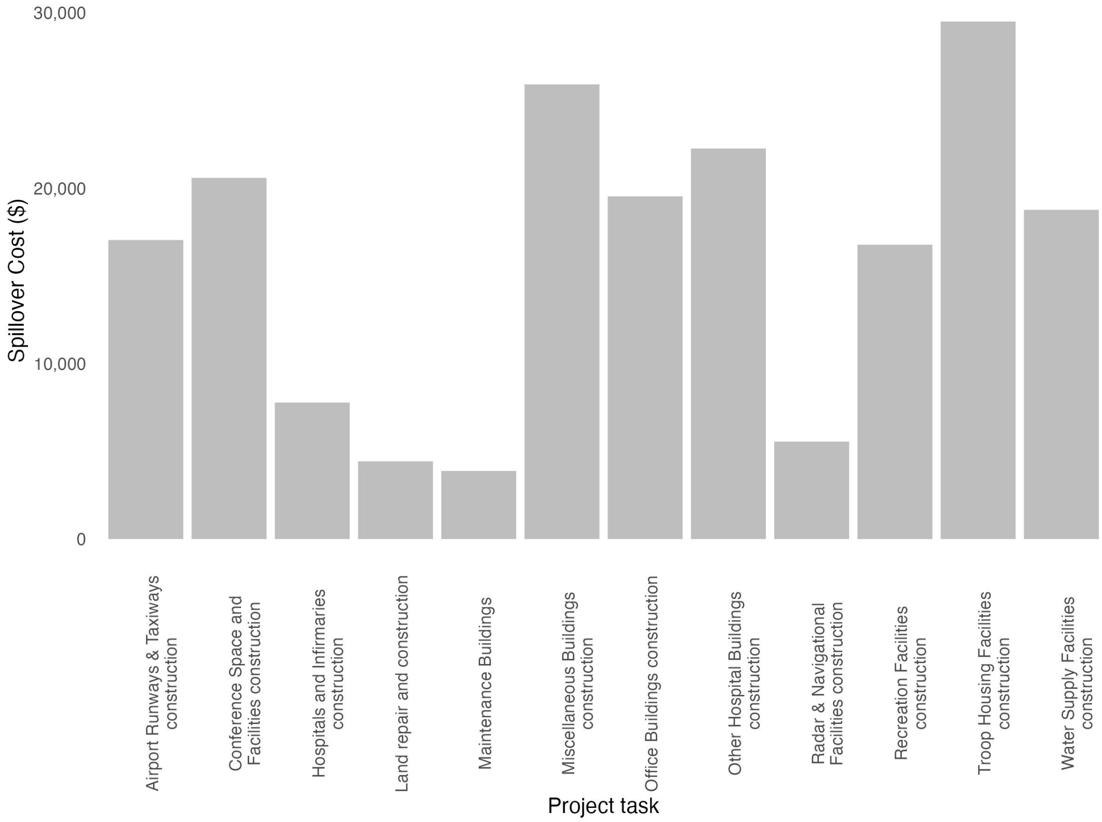
Our estimates of spillover costs using the above three methods are shown in Table B.1. Further, Figure B.1 illustrates the spillover cost for select task types in our data.

Table B.1 Expected cost of delay spillover per period

	Eichleay Formula			Modified Eichleay Formula			CD^3
	Direct costs	Indirect costs	Total Cost	Direct costs	Indirect costs	Total Cost	Total Cost
Treated projects	$33,753.11	$9,461.42	$43,214.53	$33,753.11	$24,065.65	$57,818.76	$70,778.47
Control projects	$30,327.72	$7,959.81	$38,287.53	$30,327.72	$20,344.41	$50,672.13	$60,671.65
Spillover cost			**$4,927**			**$7,146.63**	**$10,106.82**

Note. This table shows the estimated cost of delay spillovers using three different methods: Eichleay formula, Modified Eichleay formula, and CD^3 approach.

Figure B.1 Illustration of spillover costs by task type



Note. This figure illustrates the estimated cost of delay spillovers for select project tasks.

Appendix C: Non-weather disruptions

The main analysis focused on weather disruptions, given that doing so allows us to obtain cleaner results. In this section, we examine the impact of other types of disruptions in this analysis. Unlike weather-related disruptions, which can be readily identified, categorizing other disruptions is not as straightforward. Therefore, we had to establish a taxonomy. Recall that we observe a text description wherein the agency officer describes the event corresponding to a disruption. We examined these text descriptions, both manually and through word-frequency searches, to discern patterns in the disruption events. From this exercise, we identified the following frequently-occurring root causes.

1. **Bureaucratic issues:** Issues related to bureaucratic hurdles, paperwork and invoice issues, legal requirements, or policy changes that affect the project. This includes changes in building codes, permits, environmental regulations, or zoning restrictions that necessitate modifications to the project plan.

   Example: Modification to extend completion date to 30 June 2015 due to delays resolving regulatory comments on Site 7 sampling plan.

2. **Labor/personnel issues:** Issues related to labor strikes, shortage of skilled workers, or a sudden departure of key personnel that impacts the project's progress.

   Example: Add additional test fits, redesign, documentation, and construction management required to execute the revised space plan design and additional services related to the original scope of work due to labor dispute.

3. **Worksite issues:** Issues related to a malfunction or failure of machinery or equipment used in the project. This includes, for example, crane malfunction, discovery of asbestos on construction site, a generator failure, or a software glitch that affects project operations.

   Example: Date change due to asbestos- and arsenic- related issues at the building.

4. **Supply chain issues:** Issues related to shipping problems, material stockouts, or quality issues with the supplied materials.

   Example: The purpose of this modification is to a grant time extension due to late delivery and unexpected modifications the crane's runway due to the location of an electrical panel. A time extension has been granted.

We next categorize the text descriptions from the modification records into the four types of disruptions identified above. Given the size of our dataset, it is infeasible to tag all disruptions manually. Therefore, we used string matching to do a raw categorization of the data using commonly appearing strings. For instance, whenever we identified "delay due to labor," or "staff" issues, we proceeded to categorize the disruption as a labor disruption. Similarly, when we saw terms like "asbestos," "electrical", "code violation," "machine," we proceeded to tag the disruption as a worksite problem. When we saw terms like "paperwork", "permit", "administrative" issues, we classified it as a bureaucratic disruption.[1]

[1] To further validate our findings, we also conducted a secondary test via an Amazon Mechanical Turk (MTurk) task. Specifically, we asked Mturk participants to read a subsample of modification records and classify the reported disruptions in them. Although the results from this analysis were qualitatively similar, we do not report them due to the small sample size and potential for labeling inaccuracies.

Table C.1 Diff-in-diff effect by disruption type

	Bureaucratic disruptions		Worksite disruptions		Labor disruptions		Supply chain disruptions	
	I	II	III	IV	V	VI	VII	VIII
Delay days	3.03***	3.74***	6.14***	6.51***	-1.38	-1.26	6.77***	6.93***
	(1.15)	(1.14)	(1.29)	(1.28)	(6.89)	(6.77)	(1.45)	(1.40)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	451,890	451,890	339,605	339,605	97,363	97,363	400,995	400,995
R^2	0.26	0.28	0.37	0.38	0.55	0.56	0.31	0.33

Note. This table presents the estimated coefficients for diff-diff effect on delay days by the type of disruption. Each specification includes agency, contractor, task, and county fixed effects, and includes or excludes project level controls. The standard errors (reported in parentheses) are robust and clustered at the project level. Significance levels: 10% (*), 5% (**), and 1% (***)

We re-run our analysis by matching treatment and control projects using these new definitions of disruptions, and performing a diff-in-diff regression that retrieves the treatment effect by root cause—Table C.1 shows the results.

We find evidence that bureaucratic, worksite, and supply chain disruptions spill over to delay other projects. Labor disruptions, on the other hand, do not cause significant spillover effect. This finding is also consistent with our analysis that shows contractors reallocate resources from machine-intensive tasks, as opposed to labor intensive ones. While these results do provide some evidence that disruptions caused by non-weather related issues also propagate in the project network, we caution against causal interpretation of these findings due to potential for endogeneity, reporting biases, and omissions in the data (as previously discussed in §3).

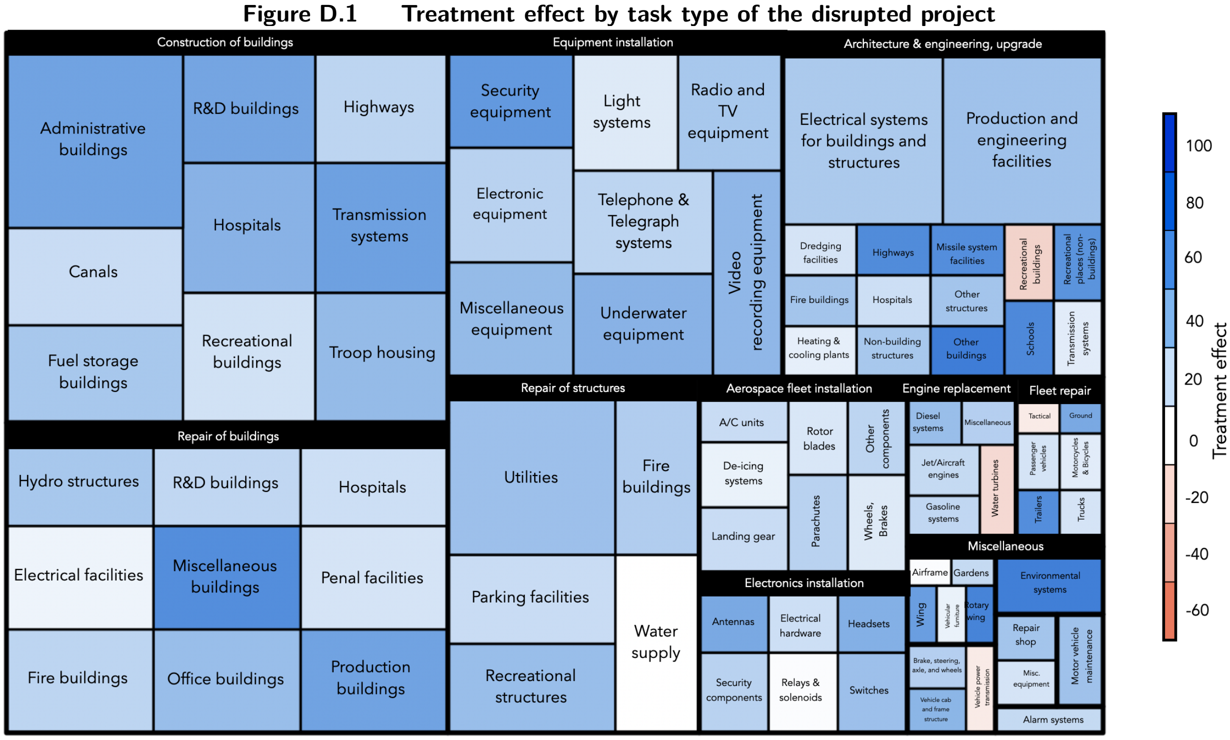
Appendix D: Additional results and summary statistics

D.1. Which tasks are more prone to reallocation externalities?

The propensity to experience a weather shock and the ease of resource reallocation may depend on the task being performed. This means that the effects studied in this paper could be highly contingent on the nature of the work. But which types of tasks are more prone to experiencing disruption externalities? To investigate this issue, our initial approach was to disaggregate the analysis and determine how the treatment effect varied as a function of the task category. That is, we obtained a treatment effect for every task category, using the 3-digit task code.

Figure D.1 displays the treatment effect by task type via a tree map, while Table D. 1 presents the five types of tasks with the largest treatment effect and the five types of tasks with the most negative treatment effect. This analysis shows allows us to see that the treatment effect is positive across most task categories, meaning that the effect is fairly generalized across the network instead of being secluded to a specific subset of tasks. From this disaggregated analysis, however, we cannot identify a clear pattern regarding which types of tasks are more prone to delays.

Figure D.1 Treatment effect by task type of the disrupted project



Note. This figure displays how the average treatment effect varies by task category of the disrupted project.

D.2. Contractor size and reallocation externalities

Our results show that localized disruptions at one project propagate to delay other projects that are connected in the project network due to self-interested resource reallocation by project participants. This prompts the question of how operational concepts such as slack time or safety capacity affect these reallocation

Table D. 1 Diff-in-Diff estimates by task type: top-5 and bottom-5 task codes

Top-5 (Three-digit task codes)

Code	Task category	Treatment effect
C12	Architecture & Engineering – Non-building structures	88.44
254	Vehicular Equipment Components – Furniture and accessories	84.73
C1C	Architecture & Engineering – Schools	81.34
C1L	Architecture & Engineering – Highways, roads, streets, bridges, and railways	78.65
Z19	Maintenance and Repair of Buildings – Miscellaneous buildings	77.56

Bottom-5 (Three-digit task codes)

Code	Task category	Treatment effect
C1F	Architecture & Engineering – Recreational buildings	-26.05
283	Engines & Turbines – Water turbines and water wheels	-22.64
Z21	Maintenance and Repair of Non-buildings – Dams	-16.70
155	Aerospace Craft And Structural Components – Space vehicles	-11.59
230	Motor Vehicles, Cycles, Trailers – Ground effect vehicles	-11.31

Table D. 2 Diff-in-diff estimates as a function of contractor size

	Treatment effect (in days)					
	I	II	III	IV	V	VI
Ln(1+Annual Revenue)	-1.32***	-1.43***	-1.30***			
	(0.16)	(0.18)	(0.18)			
Ln(1+Employees)				-2.29***	-2.58***	-2.65***
				(0.37)	(0.38)	(0.40)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Task FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	Yes	Yes	No	Yes	Yes
County FE	No	No	Yes	No	No	Yes
Num treated	25,875	25,870	25,870	25,875	25,870	25,870
R^2	0.12	0.15	0.18	0.12	0.15	0.18

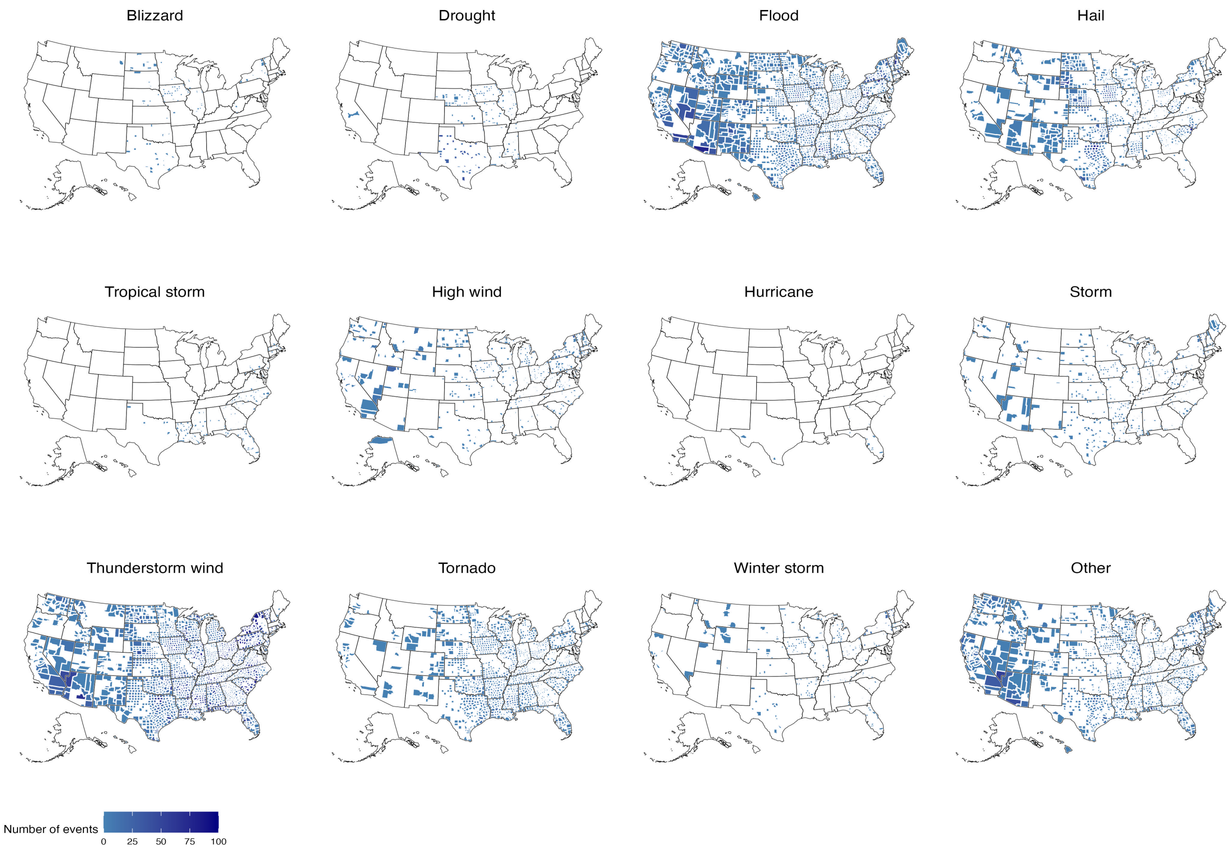
Note. This table presents the estimated coefficients for the treatment effect as a function of the contractor size. Each specification controls for the project's initial budget and duration, and number of offers received; and includes or excludes fixed effects. The standard errors (reported in parentheses) are robust and clustered at the project level. Significance levels: 10% (*), 5% (**), and 1% (***)

externalities. One could conjecture, for example, that larger contractors have a bigger resource pool and can invest in operational buffers. Therefore, projects operated by large contractors may be able to absorb the effect of a localised disruption without negative reallocation externalities. Since we do not observe the resources that contractors allocate to a project, we cannot directly observe the slack time or the level of safety capacity available. However, we can proxy for operational buffers by using two measures of contractor size in our data—the annual revenue and the number of employees. We examine how the treatment effect varies with respect to these two variables. Table D. 2 shows that, indeed, the reallocation externalities reduce as the size of the contractor increases.

D.3. Summary statistics

In this section, we present the summary statistics for the projects and weather data used in our analyses (see Figure D.2 and Tables D. 3 and D. 4), as well as on the partitioning process (in Table D. 5).

Figure D.2 Map of weather events by type



Note. Number of reported weather-related events, by county and type. Darker shades represent more events.

Table D. 3 Summary statistics: Project characteristics (drawn from the project records dataset) before matching

Variable	Type	Unit	Mean	St Dev
Project budget	Continuous	Dollars (100,000s)	5.27	11.18
Project duration	Continuous	Days	286.54	272.68
Number of bids	Count	Bids	3.45	3.71
Number of employees	Count	Employee (100s)	147.77	355.48
Annual revenue	Continuous	Dollars (in millions)	5822.71	15881.17
Competitively awarded project	Categorical	{0,1}	0.76	0.43
Cost-plus contract	Categorical	{0,1}	0.08	0.26
Fixed price contract	Categorical	{0,1}	0.88	0.33
Labor hours contract	Categorical	{0,1}	0.02	0.15
Time and Materials contract	Categorical	{0,1}	0.02	0.15
Number of projects	2,484,188			
Number of contractors	124,026			
Number of agency offices	3,559			
Number of tasks	1,188			
Number of counties	2,970			
Number of agencies	65			
Number of sub-agencies	173			
Avg. delay caused by a weather disruption	48.71 days			
Sample Timespan	Jan 01, 2011 to Sep 30, 2015			

Table D. 4 Summary statistics: Weather records

Variable	Type	Unit	Mean	St Dev
Number of deaths	Count	Person	0.010	0.330
Number of injuries	Count	Person	0.050	2.930
Property damage	Continuous	Dollars (in millions)	0.499	44.792
Event duration	Continuous	Days	1.840	6.640
Wildfire	Categorical	{0,1}	0.006	0.080
Flash flood	Categorical	{0,1}	0.060	0.240
Tornado	Categorical	{0,1}	0.020	0.140
Ice storm	Categorical	{0,1}	0.004	0.070
Drought	Categorical	{0,1}	0.060	0.230
Number of severe weather events (damage ≥$500,000)		5,808		
Time-span		Jan 1, 2011 to Dec 31, 2018		

Table D. 5 Summary statistics: Storm-hit project's network connections

Number of Connections	Percentage of projects
0	0.57
1	0.12
2	0.06
3	0.04
4	0.03
5	0.03
6	0.02
7	0.02
8	0.02
9	0.01
≥10	0.14

Note. This table shows the number of network connections that a given storm-hit project concurrently had at a given time (not the sum of connections over the entire project duration).

Table D. 6 Partitions in Step 1

Partition size	Number of partitions
1	272,403
2	58,876
3	26,327
4	14,873
5	9,879
6	6,875
7	5,026
8	3,924
9	3,116
≥10	27,788

Table D. 7 Period specific partitions in Step 3

Partition size	Number of partitions	% Treated
1	986,666	19.85
2	175,822	21.17
3	70,199	21.90
4	37,204	22.40
5	22,758	22.86
6	15,527	22.77
7	10,848	22.85
8	8,197	23.07
9	6,194	24.47
≥ 10	41,489	29.48

Table D. 8 Identifiable partitions in Step 4

Partition size	Number of partitions
1	0
2	17,644
3	10,581
4	7,010
5	4,899
6	3,703
7	2,862
8	2,250
9	1,910
10	16,424

References

Autor, David H. 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of labor economics* **21**(1) 1–42.

Beaty, Curtis, David Ellis, Brianne Glover, Bill Stockton, et al. 2016. Assessing the costs attributed to project delay during project pre-construction stages. Tech. rep., Texas. Dept. of Transportation.

Bertrand, Marianne, Esther Duflo, Sendhil Mullainathan. 2004. How much should we trust differences-in-differences estimates? *The Quarterly journal of economics* **119**(1) 249–275.

Calvo, Eduard, Ruomeng Cui, Juan Camilo Serpa. 2019. Oversight and efficiency in public projects: A regression discontinuity analysis. *Management Science* **65**(12) 5651–5675.

Cameron, A Colin, Douglas L Miller. 2015. A practitioner's guide to cluster-robust inference. *Journal of Human Resources* **50**(2) 317–372.

Keogh, Geoffrey, Alan W Evans. 1992. The private and social costs of planning delay. *Urban Studies* **29**(5) 687–699.

Lewis, Gregory, Patrick Bajari. 2014. Moral Hazard, Incentive Contracts, and Risk: Evidence from Procurement. *The Review of Economic Studies* **81**(3) 1201–1228.

Monfared, Sam, Andrey Pavlov. 2017. Political risk affects real estate markets. *The Journal of Real Estate Finance and Economics* 1–20.

Stuart, Elizabeth A, Brian K Lee, Finbarr P Leacy. 2013. Prognostic score–based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of clinical epidemiology* **66**(8) S84–S90.