Time-dependent treatment effect

0. Descriptive statistics of data and summary

We have 10 quarters in the observation horizon, which means that a project may report a delay (or expedition) in at most 10 quarters. The following table shows that **the majority of projects (> 75%) in our data report only one delay/expedition over 10 quarters**.

This feature of our data implies that although we aggregate data on a quarterly basis, our estimates should not be interpreted as "probability of delay/expedition per quarter" and "avg. delay per quarter" of a given project. Aggregation over multiple quarters won't do much (as is manifested in the regression results).

Num. of quarters with delay or expedition	Num. of projects	Fraction of all projects	Cumu. fraction	Num. of small projects	Fraction of small projects	Cumu. fraction	Num. of large projects	Fraction of large projects	Cumu. fraction
1	17,679	75.8%	75.80%	8,667	76.92%	76.92%	9,012	74.74%	74.74%
2	4,132	17.71%	93.51%	1,962	17.41%	94.34%	2,170	18.00%	92.74%
3	1,078	4.62%	98.13%	473	4.20%	98.54%	605	5.02%	97.76%
4	324	1.39%	99.52%	128	1.14%	99.67%	196	1.62%	99.39%
5	79	0.34%	99.86%	23	0.20%	99.87%	56	0.46%	99.85%
6	20	0.09%	99.95%	11	0.010%	99.97%	9	0.07%	99.92%
7	8	0.03%	99.98%	1	0.009%	99.98%	7	0.058%	99.98%
8	3	0.013%	99.995%	2	0.02%	100%	1	0.008%	99.99%
9	1	0.004%	100%	0	0	100%	1	0.008%	100%
Total num. of projects	23,324			11,267			12,057		

A few questions to think about:

- How to interpret the estimates from our model using the entire data set?
- Why are we using quarterly aggregated data? Enlarging the sample size is a good reason. In the logistic regressions, for example, we no longer have enough statistics power to have significant estimates after aggregating on the year level (four quarter). But how to interpret the result with quarterly data?
- What do we want to estimate? Shall we focus on a subset of projects, e.g., the ones with only one delay over the entire observation horizon, to get a cleaner understanding?

1. Time-dependent treatment effect

1.1. Linear regression on nonzero delays

Significance code: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. All errors clustered at PSC level.

Table 1. Estimation results for nonzero delay with projects that delay only once					Table 2. Estimation results for nonzero delay with all projects			
	(1)	(2)	(3)			(1)	(2)	(3)
	Estimate (st. error)	Estimate (st. error)	Estimate (st. error)			Estimate (st. error)	Estimate (st. error)	Estimate (st. error)
Treat	-19***	-18***	-13**		Treat	-21***	-21***	-11**
	(5)	(5)	(5)			(5)	(5)	(4)
Treat*Post	30***	30***	30***		Treat*Post	29***	29***	26***
	(10)	(11)	(10)			(8)	(8)	(7)
Treat*Post_Qrtr2	-16	-16	-17		Treat*Post_Qrtr2	-14*	-14*	-14*
	(12)	(12)	(10)			(8)	(8)	(7)
Treat*Post_Qrtr3	-30**	-30**	-18*		Treat*Post_Qrtr3	-20**	-20	-9
	(15)	(15)	(10)			(12)	(13)	(7)
Treat*Post_Qrtr4	-6	-6	-11		Treat*Post_Qrtr4	-5	-5	-5
	(10)	(10)	(10)			(8)	(8)	(7)
Treat*Post_Qrtr5	-4	-5	-6		Treat*Post_Qrtr5	-4	-4	-2
	(10)	(10)	(10)			(6)	(6)	(6)
Time	No	Yes	Yes		Time	No	Yes	Yes
PSC, NAICS	No	No	Yes		PSC, NAICS	No	No	Yes
Adjusted R-squared	0.03	0.03	0.15		Adjusted R-square	0.01	0.02	0.15
N	17,679	17,679	17,679		N	23,324	23,324	23,324

Table 1 shows the results with projects that delay only once. Interpretation of the coefficients are as follows:

- *Treat*Post*: The baseline treatment effect in all five quarters after QP. If a small-business project delays *any time within the five quarters after QP*, then the launch of QP makes the project delays 30 days more than it would before QP.
- *Treat*Post_Qrtr2*: The additional treatment effect two quarters after QP launch. Insignificant estimate means a project that delays two quarters after QP behaves the same way as the baseline.
- *Treat*Post_Qrtr3*: The additional treatment effect three quarters after QP launch. If a small-business project delays *three quarters after QP*, then the launch of QP makes the project delays 18 days less than the baseline, i.e., it delays 12 days more than without QP.

Table 2 shows the results with all projects. I am not sure how to interpret the estimates...

1.2. Logistic regression on delay probability

Significance code: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. All errors clustered at PSC level.

Table 3. Estimation results for delay probability with projects that delay only once					Table 4. Estimation results for delay probability with all projects			
	(1)	(2)	(3)			(1)	(2)	(3)
	Estimate (st. error)	Estimate (st. error)	Estimate (st. error)			Estimate (st. error)	Estimate (st. error)	Estimate (st. error)
Treat	-0.7*** (0.08)	0.06 (0.08)	0.05 (0.06)		Treat	-1.4*** (0.07)	-0.1* (0.06)	-0.18*** (0.05)
Treat*Post	0.4*** (0.1)	-0.34*** (0.10)	-0.30*** (0.10)		Treat*Post	1.3*** (0.1)	-0.01 (0.07)	0.00 (0.07)
Treat*Post_Qrtr2	0.3** (0.1)	0.3** (0.1)	0.3** (0.1)		Treat*Post_Qrtr2	0.04 (0.08)	0.04 (0.08)	0.06 (0.09)
Treat*Post_Qrtr3	0.3* (0.2)	0.3* (0.2)	0.2 (0.2)		Treat*Post_Qrtr3	0.06 (0.1)	0.06 (0.1)	0.06 (0.1)
Treat*Post_Qrtr4	-0.4*** (0.1)	0.4*** (0.1)	0.4*** (0.1)		Treat*Post_Qrtr4	0.15** (0.07)	0.15** (0.07)	0.18** (0.07)
Treat*Post_Qrtr5	-0.5*** (0.1)	0.5*** (0.1)	0.5*** (0.1)		Treat*Post_Qrtr5	0.20** (0.08)	0.20** (0.1)	0.22*** (0.08)
Time	No	Yes	Yes		Time	No	Yes	Yes
PSC, NAICS	No	No	Yes		PSC, NAICS	No	No	Yes
N	45,688	45,688	45,424		N	125,426	125,426	124,297

Table 3 shows logistic regression results with projects that delay or expedite only once. Interpretation of the coefficients are as follows:

- *Treat*Post*: The baseline treatment effect in all five quarters after QP.
    - A small-business project is less likely to delay after QP in the first and third quarter after QP.
- *Treat*Post_Qrtr2*: The additional treatment effect two quarters after QP launch over the baseline. A project has about the same likelihood to delay in the second quarter after QP as it would without QP.

- *Treat*Post_Qrtr4*: The additional treatment effect four quarters after QP launch. A project is more likely to delay four quarters after QP launch. The ratio of delay odds before and after QP is $e^{-0.3+0.4} = 1.10$.

- *Treat*Post_Qrtr5*: The additional treatment effect five quarters after QP launch. A project is more likely to delay five quarters after QP launch. The ratio of delay odds before and after QP is $e^{-0.3+0.5} = 1.22$.

Table 4 shows logistic regression results with all projects. I am not sure how to interpret them...

2. Results with aggregated data

The earlier approach of aggregating data is problematic, as the delay in the same quarter is counted multiple times, thus creating correlation among the dependent variables and inflating results.

I redid the aggregation on half-year and yearly basis. On both aggregation levels, I consider one year before and one year after QP. This means that

- On the half-year (bi-quarter) aggregation level, we have two bi-quarter observations before QP and two bi-quarter observations after it.
- On the year (quad-quarter) aggregation level, we have one observation before QP and one after it.

Overall results:

- Nonzero delays: No significant change in the treatment effect Treat*Post using quarterly, bi-quarterly, and quad-quarterly delay data. (Expected given that the majority of projects delay only once in the entire observation horizon.)
- Probability of delay: Point estimate of the treatment effect Treat*Post have the same sign using quarterly, bi-quarterly, and quad-quarterly delay data but the significance level drops. In quad-quarterly delay data, the result is insignificant. Possibly due to the reduction in sample size.