Green Infrastructure: The Effects of Urban Rail Transit on Air Quality Supplemental Appendix

Yihsu Chen
University of California, Merced

Alexander Whalley
University of California, Merced & NBER

February 1, 2011

1 Supplementary Appendix

Table A1 reports the estimates from the alternative specifications of the DB-OLS equation (2) analysis for the three transportation-related pollutants. We examine the robustness of our results along three dimensions: reporting, polynomial order specification, and covariate choice.

First, we examine whether our results are sensitive to the sample used in Panel A of Table A1. This is an important potential concern as our data are subject to some non-reporting, particularly for NO_x pollution. We check for whether reporting bias is a key concern by limiting the sample to high reporting hours and stations, excluding outliers, including station fixed effects. For our high reporting sample we include only those hours or stations where the fraction of missing values for NO_x pollution is less than 10 percent.[1] For our trimmed sample we exclude observations for the relevant pollutant outside of the middle 98 percentiles of the pollutant distribution. The results of estimating our central specification with the limited sample are presented in panel A. We find little evidence that the magnitude of the estimated effect substantially changes with the change in the sample. Both the magnitude of the estimates are weakened when the outliers are dropped. However, the sign of the point estimate remains negative and statistically significant. The precision of the point estimates is also weaker with the station fixed effects specification, but the magnitude of the point estimates are quite similar to the baseline models.

We next examine whether our results are sensitive to the order of the specification of the polynomial time trend. To do so we repeat our regression discontinuity analyses with 1st- and 5th-order time trend polynomials. The results are summarized in the Panel B in Table A1. With the first order polynomial the magnitude of the point estimate is somewhat smaller than the baseline estimate (though it remains above the difference-in-difference estimate) and imprecisely estimated. The point estimate with the 5th order polynomial remains negative and though somewhat smaller in magnitude to the baseline estimate, however is remains statistically significant at the 5% level. We then examine how sensitive our results are to the method used to account for serial correlation in the

[1]This restriction excludes two hours (4AM and 7AM) and three stations (Guting, Wanli, Yangmin) from the sample respectively.

errors. In Panel C of Table 6 we report the results of estimation for the model with Newey and West standard errors (Newey and West, 1987), which allow for the correlation across groups with a five-week lag. The Newey and West standard errors are close in magnitude to the standard errors that allow arbitrary correlation in five-week group in the baseline models.[2]

Lastly, we examine how sensitive our results are to the inclusion or exclusion of various controls. We consider the effects of three sets of controls. We first look at the effects of including gas prices. While high frequency gas prices are not available in Taiwan during our sample period, we are able to use daily gas prices in Singapore to measure whether international trends in the cost of fuel can account for our main results. We find little evidence that the Taipei Metro effects are sensitive to the inclusion of gas prices. Next, while we seek to estimate the effect of the opening of the Metro alone, whether the inclusion of gas content regulation events can account for our results is well worth examining. We find little evidence that our results are sensitive to the exclusion of these controls, as the negative effect remains statistically significant at the 10% level. As noted in the main text the humidity variable is not smooth at the discontinuity and is also subject to some non-reporting. We lastly examine whether our results are sensitive to the inclusion of the humidity covariates. Again, we find little evidence that our results are highly sensitive to this set of controls, as the estimates remain statistically significant at the 10% level.

[2]The magnitudes of the estimates differ slightly from those in the baseline. This difference arises because these models are estimated at the day level rather then hour level as in the baseline models.

TABLE A1: The Effect of Metro Opening on Transportation Source Pollutants, Alternative Samples and Specifications: Ridership Discontinuity Based OLS

Dependent Variable:	Log(CO)	Log(NO_x)	Log(O_3)
Model:	DB-OLS	DB-OLS	DB-OLS
	(1)	(2)	(3)
Baseline	-0.156**	-0.083	-0.037
	(0.059)	(0.052)	(0.063)
Panel A: Alternative Samples			
High Reporting Stations Only	-0.153**	-0.132**	0.015
	(0.059)	(0.066)	(0.093)
High Reporting Hours Only	-0.155**	-0.081**	-0.034
	(0.049)	(0.054)	(0.064)
Excluding Outliers	-0.118**	-0.056	-0.018
	(0.056)	(0.045)	(0.057)
Station Fixed Effects	-0.154*	-0.107	-0.008
	(0.081)	(0.082)	(0.082)
Panel B: Alternative Higher Order Polynomial Specifications			
First – Order Polynomial	-0.066	-0.093**	-0.052
	(0.051)	(0.042)	(0.099)
Fifth – Order Polynomial	-0.073**	-0.099**	0.011
	(0.036)	(0.045)	(0.124)
Panel C: Alternative Specifications			
Newey-West Standard Errors	-0.136**	-0.059	-0.164*
	(0.053)	(0.052)	(0.093)
Including Singapore Gas Price	-0.184**	-0.084**	0.117
Covariate	(0.056)	(0.066)	(0.095)
Excluding Gas Content Regulation	-0.101*	-0.059	-0.008
Covariates	(0.058)	(0.068)	(0.065)
Excluding Humidity Covariates	-0.100*	-0.048	-0.031
	(0.053)	(0.065)	(0.074)

Notes: Source: Authors' Calculations from Taiwanese EPA air quality monitoring network data. Each cell reports the result from one regression with controls for a fifth-order polynomial time trend, gas content regulation events, quartics in wind speed, temperature, and humidity, as well as, one hour lags of these variables unless otherwise indicated. The unit of observation is hour. The sample is all observations for one full year before and after the TM opening date, unless otherwise indicated. The main entries in columns (1) - (3) report the coefficient estimate from fitting model (2) in the text by ordinary least squares, with the standard errors clustered at the 5-week level reported in brackets, unless otherwise indicated. See the text for the precise details of the specification in each row of the column. *** indicates significantly

different from zero at the 1% level. ** indicates significantly different from zero at the 5% level. * indicates significantly different from zero at the 10% level.

TABLE A2: Pollution and Climatic Conditions Before Taipei Metro Opening, by City

	Taipei	East Coast	Kaohsiung	(1) – (2) t-stat	(1) – (3) t-stat
	(1)	(2)	(3)	(4)	(5)
A. Primarily Transportation Source Pollutants					
Carbon Monoxide (CO)	0.834	0.580	0.827	-45.53	-1.22
	[0.440]	[0.262]	[0.324]	(0.000)	(0.223)
	N=8735	N=8326	N=8375		
Nitrogen Oxides (NO_x)	0.035	0.013	0.043	-86.90	24.79
	[0.021]	[0.009]	[0.024]	(0.000)	(0.000)
	N=8437	N=8053	N=8415		
B. Indirect Transportation Source Pollutants					
Ground-Level Ozone	0.023	0.022	0.021	-2.94	-8.21
(O_3)	[0.015]	[0.012]	[0.023]	(0.003)	(0.000)
	N=8731	N=8503	N=8731		
C. Primarily Non-Transportation Source Pollutants					
Particulate Matter	48.868	36.551	88.235	-31.48	65.55
(PM10)	[27.898]	[23.309]	[48.705]	(0.000)	(0.000)
	N=8734	N=8566	N=8734		
Sulfur Dioxide (SO_2)	0.006	0.001	0.014	-94.48	76.29
	[0.005]	[0.015]	[0.009]	(0.000)	(0.000)
	N=8735	N=7439	N=8735		
D. Weather					
Wind Speed	2.23	2.18	1.97	-3.10	-14.52
	[1.16]	[1.16]	[1.17]	(0.002)	(0.000)
	N=8736	N=8627	N=8735		
Temperature	21.68	24.11	24.67	30.12	35.16
	[6.12]	[4.39]	[5.10]	(0.000)	(0.000)
	N=8735	N=8641	N=8735		
Humidity	73.23	70.11	73.08	-20.35	-1.17
	[8.22]	[11.28]	[8.22]	(0.000)	(0.241)
	N=8371	N=8273	N=8368		

Notes: Source: Authors' Calculations from Taiwanese EPA air quality monitoring network data. The unit of observation is hour for all variables in panels A-D. All pollutants are expressed in parts per million, wind speed is expressed in meters per second, temperature is expressed in degrees Celsius, and humidity is expressed in percentage terms. The main entries in columns (1), (2) and (3) report the mean level of the variable indicated in the row heading and the sample indicated in the column heading. The entries in square brackets in columns (1), (2) and (3) report the standard deviation of the variable indicated in the row heading and the sample indicated in the column heading. The t-statistic and the p-value in square brackets

for the hypothesis test that the variable indicated in the row heading does not differ between columns (2) and (3) is reported in column (4).