

Thesis Progress

11/3/23

Original Idea: Taiwan HSR

Treatment: Opening of HSR lines

Methodology: DiD

Shortcomings: Lack observations in terms of geography, since the line opened all at once rather than having a staggered opening

New focus: Mass Rapid Transit System

Why?

- Data availability: geographical abstraction in 2007
- MRT opened in 1996

Methodology: Event study

- Comparing households outcomes in districts that were connected in year 1, 2, 3, etc or not at all

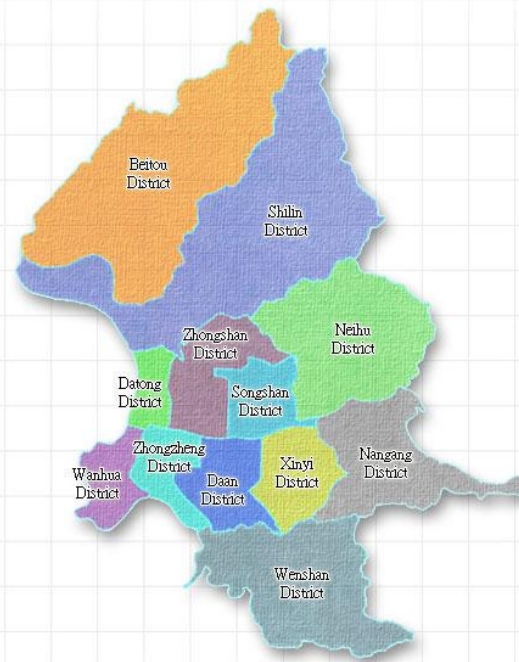
Outcomes of interest:

- household expenditure on public transit, (in survey data I currently have)
- Rent prices (survey data)
- occupation/industry, (survey data)
- sale of personal vehicles (available on CEIC)
- Tourism (CEIC)

Data and granularity

Taipei has 12 districts

- I can look at outcomes in each district for over 15 years at least, most likely more, so a size of at least 180 ($12 * 15$)
- Data collection: Documented districts that were connected by year, as well as the number of stations that exist within each district at a given year



Current data set

37,500 observations, appended survey data years 1991-2006, selected those only living in Taipei as it was a national survey data set

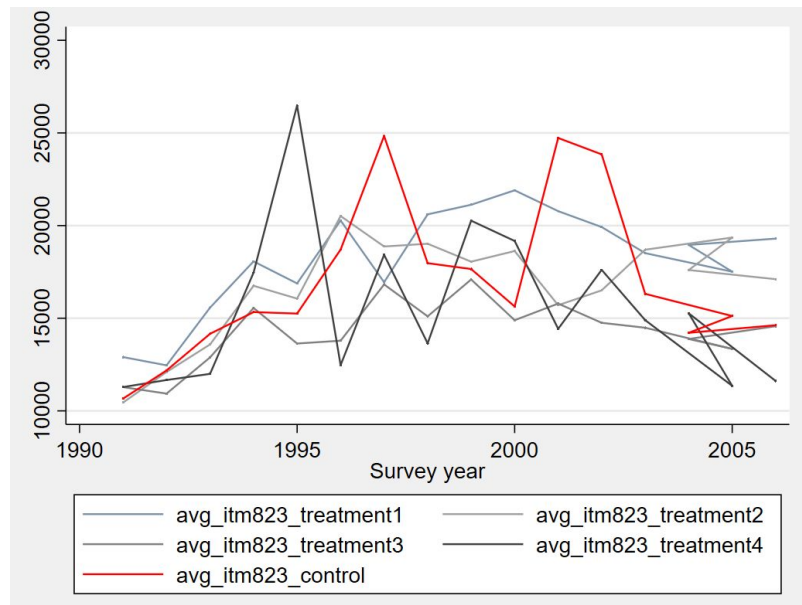
~3,125 obs for each district, each equally represented

Collected data on when stations opened up, created 4 different treatment groups and 1 control group

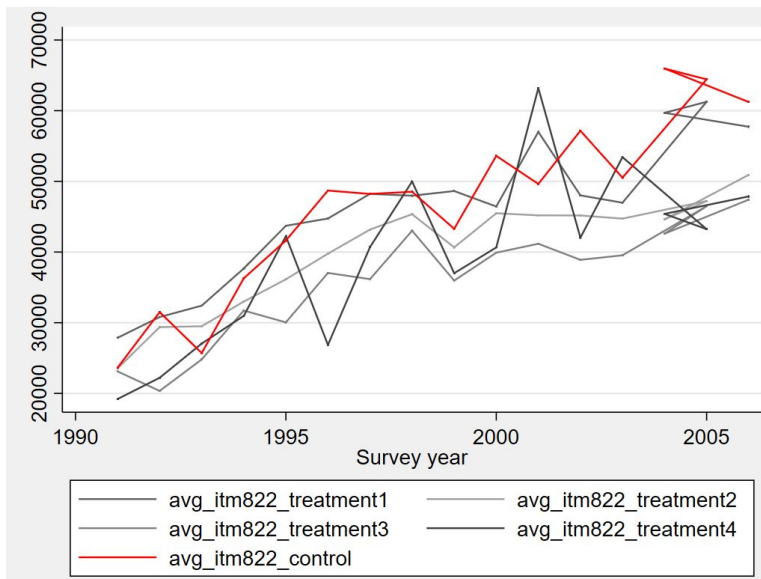
```
gen treatment_group1 = 1 if inlist(area, 31, 33, 38)
gen treatment_group2 = 1 if inlist(area, 36, 35, 34, 41, 42)
gen treatment_group3 = 1 if inlist(area, 32, 37)
gen treatment_group4 = 1 if inlist(area, 39)
gen control_group = 1 if inlist(area, 40)
```

Trends over time

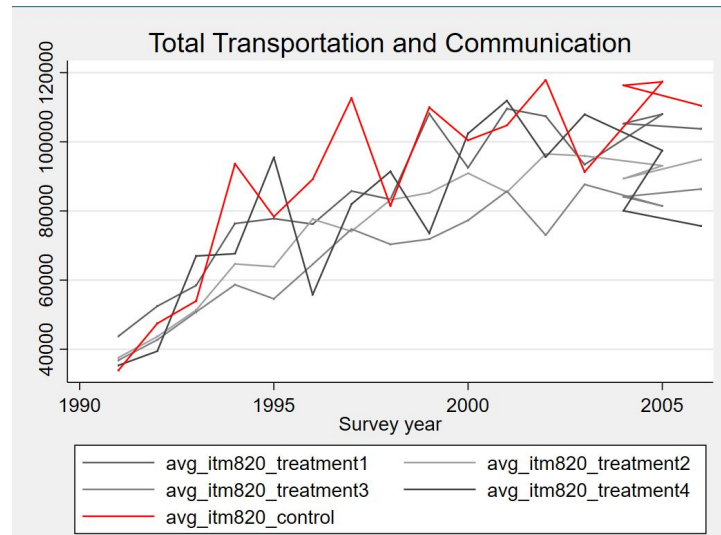
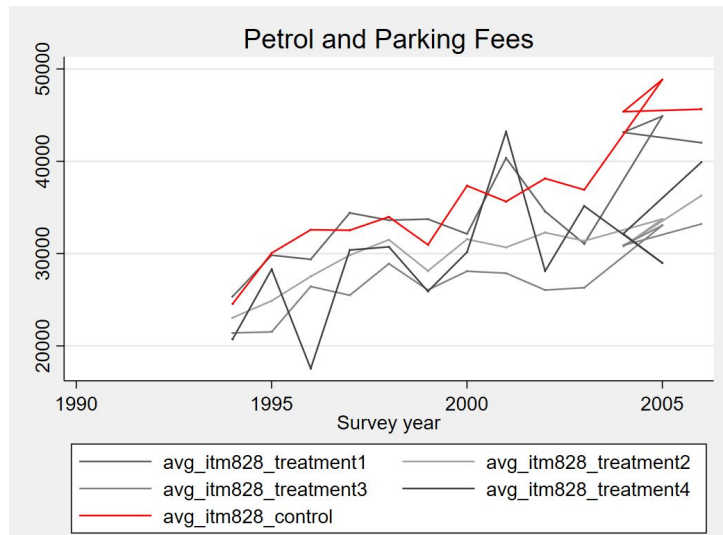
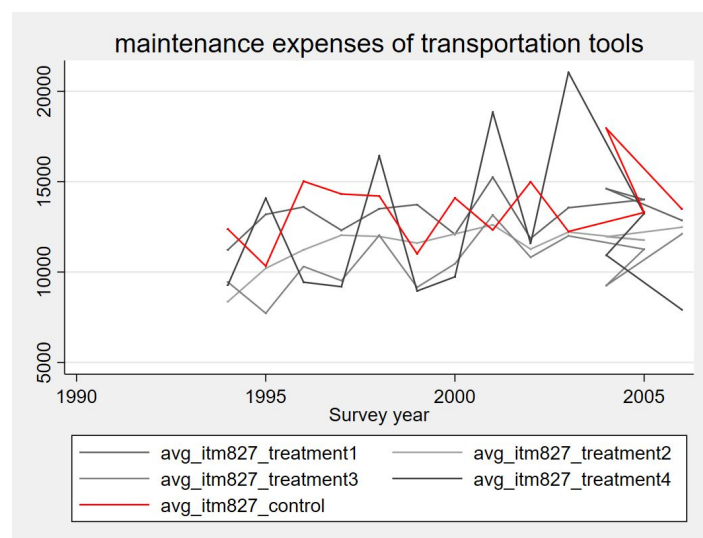
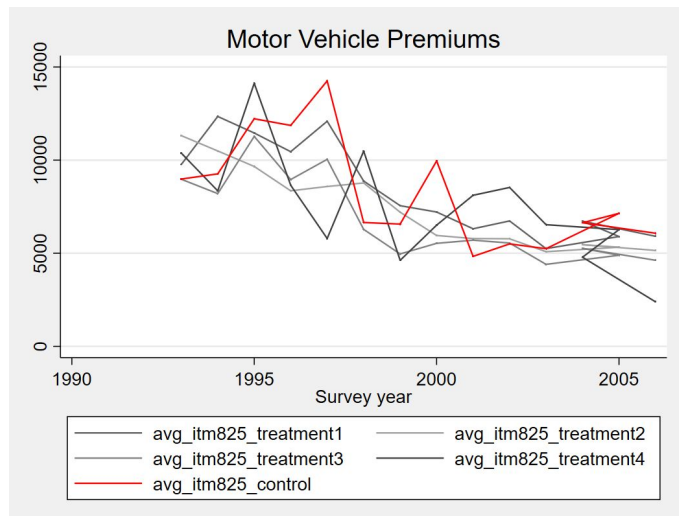
Itm 823: “purchase transportation”



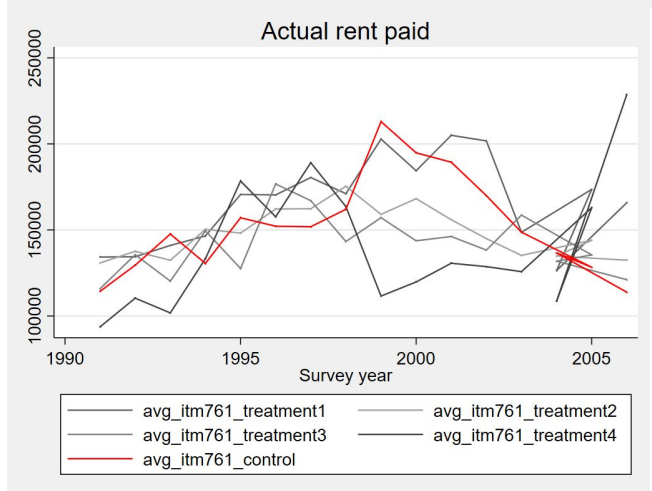
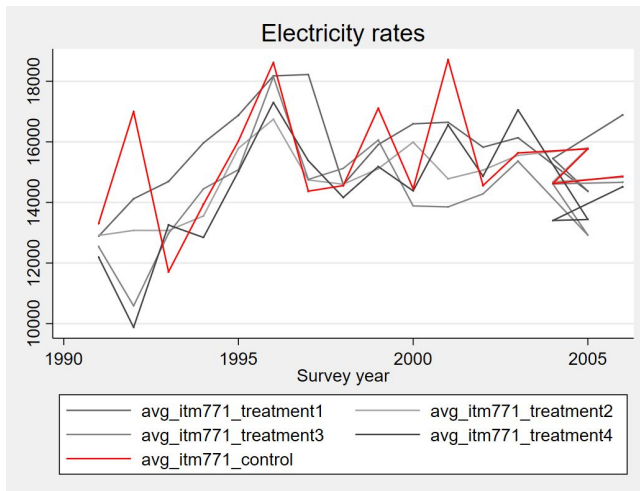
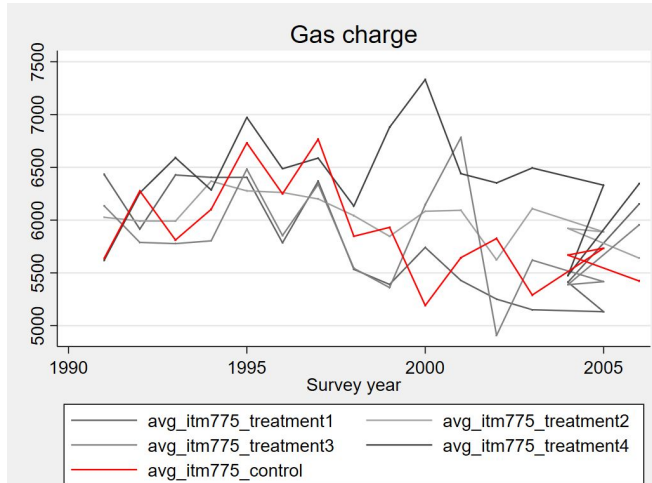
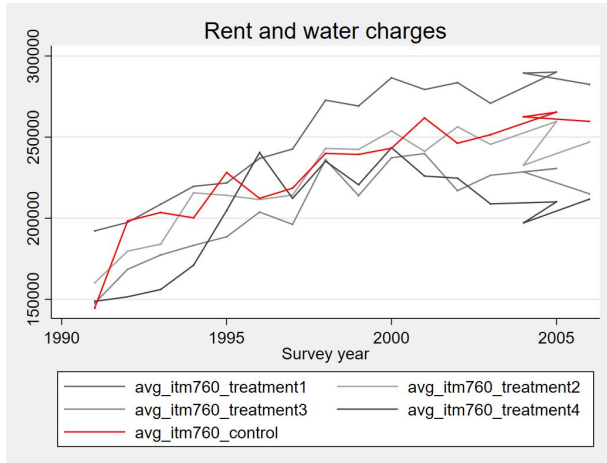
Itm 822: “expenses on maintenance of personal transportation equipment”



```
foreach var in itm820 itm821 itm822 itm823 itm825 itm826 itm827 itm828 {  
    local varname "`var'"  
    egen avg_`var'_treatment1 = mean(`var') if treatment_group1 == 1, by(year)  
    egen avg_`var'_treatment2 = mean(`var') if treatment_group2 == 1, by(year)  
    egen avg_`var'_treatment3 = mean(`var') if treatment_group3 == 1, by(year)  
    egen avg_`var'_treatment4 = mean(`var') if treatment_group4 == 1, by(year)  
    egen avg_`var'_control = mean(`var') if control_group == 1, by(year)  
}
```



Other outcomes



Simple OLS on outcomes (itm 823: purchase transit)

Explanatory variable: num_stations, the number of stations that exist within the observation's (household's) district

Controls: year, income, gender, household head's age, education, employment status, marital status, industry, and employment classification by working place

itm823	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	170.6478	44.42298	3.84	0.000	83.57684	257.7188
year	118.3208	28.26217	4.19	0.000	62.92572	173.7159
itm190	.0064527	.0001762	36.61	0.000	.0061072	.0067981
b3_1	2124.634	266.2129	7.98	0.000	1602.844	2646.423
b4_1	151.9091	10.64318	14.27	0.000	131.048	172.7702
b5_1	196.3703	61.59188	3.19	0.001	75.64748	317.0931
b11_1	-991.9979	99.6357	-9.96	0.000	-1187.288	-796.7076
b13_1	-47.73969	616.1295	-0.08	0.938	-1255.381	1159.901
b14_1	2017.003	485.272	4.16	0.000	1065.849	2968.158
b16_1	-9.829299	3.176413	-3.09	0.002	-16.05521	-3.60339
_cons	-231531.1	56342.23	-4.11	0.000	-341964.4	-121097.8

On Industry

```
. regress transit_industry num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	29,850
				F(10, 29839)	=	46.40
Model	14.1126555	10	1.41126555	Prob > F	=	0.0000
Residual	907.525401	29,839	.030414069	R-squared	=	0.0153
				Adj R-squared	=	0.0150
Total	921.638057	29,849	.030876681	Root MSE	=	.1744

61: transport, storage & communication

transit_in~y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	.0000763	.0004318	0.18	0.860	-.00077	.0009226
itml90	6.19e-09	1.71e-09	3.61	0.000	2.83e-09	9.54e-09
year	-.0010254	.0002745	-3.74	0.000	-.0015634	-.0004875
b3_1	-.0181452	.0025833	-7.02	0.000	-.0232086	-.0130818
b4_1	-.0000218	.0001034	-0.21	0.833	-.0002244	.0001809
b5_1	-.0032667	.0005984	-5.46	0.000	-.0044396	-.0020939
b11_1	-.0019194	.000968	-1.98	0.047	-.0038168	-.000022
b13_1	-.0333056	.0059876	-5.56	0.000	-.0450415	-.0215697
b14_1	-.0017207	.0047121	-0.37	0.715	-.0109566	.0075152
b16_1	.0000365	.0000308	1.19	0.236	-.0000238	.0000968
_cons	2.17511	.5471527	3.98	0.000	1.102667	3.247553

Industry 2

```
regress public_industry num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	29,850
Model	83.6692452	10	8.36692452	F(10, 29839)	=	167.33
Residual	1492.01213	29,839	.050002082	Prob > F	=	0.0000
				R-squared	=	0.0531
				Adj R-squared	=	0.0528
Total	1575.68137	29,849	.052788414	Root MSE	=	.22361

91: public administration

public_ind~y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	-.0004717	.0005536	-0.85	0.394	-.0015569	.0006134
itml90	1.84e-09	2.20e-09	0.84	0.403	-2.47e-09	6.14e-09
year	-.0068314	.0003519	-19.41	0.000	-.0075212	-.0061416
b3_1	.0192787	.0033123	5.82	0.000	.0127864	.025771
b4_1	.0014141	.0001325	10.67	0.000	.0011543	.0016739
b5_1	.0116547	.0007672	15.19	0.000	.0101509	.0131586
b11_1	.0195581	.0012412	15.76	0.000	.0171253	.021991
b13_1	-.1584252	.0076773	-20.64	0.000	-.173473	-.1433774
b14_1	.0156259	.0060419	2.59	0.010	.0037836	.0274682
b16_1	-.0000877	.0000395	-2.22	0.026	-.000165	-.0000103
_cons	13.58139	.7015608	19.36	0.000	12.2063	14.95648

Industry 3

```
. regress social_industry num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	29,850
Model	119.471608	10	11.9471608	F(10, 29839)	=	155.09
Residual	2298.55801	29,839	.077032005	Prob > F	=	0.0000
				R-squared	=	0.0494
				Adj R-squared	=	0.0491
Total	2418.02961	29,849	.081008731	Root MSE	=	.27755

81: social & personal service

social_ind~y	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	.0048074	.0006872	7.00	0.000	.0034605	.0061542
itml90	6.77e-09	2.73e-09	2.48	0.013	1.43e-09	1.21e-08
year	-.0046334	.0004368	-10.61	0.000	-.0054895	-.0037772
b3_1	.0629857	.0041113	15.32	0.000	.0549275	.071044
b4_1	.0015585	.0001645	9.47	0.000	.001236	.0018809
b5_1	.0065317	.0009523	6.86	0.000	.0046652	.0083983
b11_1	.0046382	.0015406	3.01	0.003	.0016185	.0076578
b13_1	-.1608975	.009529	-16.88	0.000	-.1795749	-.1422202
b14_1	-.0044278	.0074992	-0.59	0.555	-.0191265	.0102708
b16_1	.0000527	.000049	1.08	0.282	-.0000433	.0001487
_cons	9.307761	.8707769	10.69	0.000	7.601	11.01452

Industry 4

```
regress public_sector num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	29,850
				F(10, 29839)	=	412.47
Model	360.801911	10	36.0801911	Prob > F	=	0.0000
Residual	2610.13223	29,839	.087473851	R-squared	=	0.1214
				Adj R-squared	=	0.1211
Total	2970.93414	29,849	.099532116	Root MSE	=	.29576

public_sec~r	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
um_stations	.0011641	.0007323	1.59	0.112	-.0002712	.0025993
itml90	1.87e-11	2.91e-09	0.01	0.995	-5.68e-09	5.71e-09
year	-.006499	.0004655	-13.96	0.000	-.0074114	-.0055867
b3_1	.0531181	.0043811	12.12	0.000	.044531	.0617051
b4_1	.0034541	.0001753	19.70	0.000	.0031105	.0037978
b5_1	.0326151	.0010148	32.14	0.000	.0306261	.0346042
b11_1	.0589753	.0016417	35.92	0.000	.0557575	.0621932
b13_1	-.4140957	.0101544	-40.78	0.000	-.4339987	-.3941927
b14_1	.0551961	.0079913	6.91	0.000	.0395329	.0708594
b16_1	-.0002786	.0000522	-5.34	0.000	-.0003809	-.0001763
_cons	12.69493	.9279197	13.68	0.000	10.87617	14.51369

by sector of activity

1: public enterprises

2: general governments

3: others

Other outcomes

Motor premiums

```
. regress itm825 num_stations itm190 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	17,495
Model	2.2933e+11	10	2.2933e+10	F(10, 17484)	=	192.53
Residual	2.0826e+12	17,484	119114245	Prob > F	=	0.0000
				R-squared	=	0.0992
				Adj R-squared	=	0.0987
Total	2.3119e+12	17,494	132155171	Root MSE	=	10914

itm825	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	-96.46134	31.01182	-3.11	0.002	-157.2476	-35.67509
itm190	.0038969	.0001304	29.89	0.000	.0036413	.0041525
year	-468.1944	24.22442	-19.33	0.000	-515.6767	-420.7122
b3_1	348.7622	208.1146	1.68	0.094	-59.16313	756.6876
b4_1	-9.442485	8.897988	-1.06	0.289	-26.88343	7.998459
b5_1	237.3089	51.2877	4.63	0.000	136.7799	337.8379
b11_1	-1176.669	77.35225	-15.21	0.000	-1328.287	-1025.051
b13_1	4065.521	513.5795	7.92	0.000	3058.854	5072.188
b14_1	452.2899	426.0417	1.06	0.288	-382.7942	1287.374
b16_1	-3.774923	2.609958	-1.45	0.148	-8.890702	1.340855
_cons	941316.8	48354.11	19.47	0.000	846537.9	1036096

Transport and communication tools

```
. regress itm826 num_stations itm190 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1
```

Source	SS	df	MS	Number of obs	=	2,969
Model	1.0578e+13	10	1.0578e+12	F(10, 2958)	=	17.25
Residual	1.8134e+14	2,958	6.1306e+10	Prob > F	=	0.0000
				R-squared	=	0.0551
				Adj R-squared	=	0.0519
Total	1.9192e+14	2,968	6.4664e+10	Root MSE	=	2.5e+05

itm826	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	-2948.227	1621.477	-1.82	0.069	-6127.565	231.1111
itm190	.0568096	.0064457	8.81	0.000	.0441711	.0694482
year	-11797.42	1433.908	-8.23	0.000	-14608.97	-8985.857
b3_1	-8499.389	11187.39	-0.76	0.447	-30435.24	13436.46
b4_1	37.55217	485.3767	0.08	0.938	-914.1582	989.2625
b5_1	1363.497	2828.755	0.48	0.630	-4183.03	6910.024
b11_1	-11055.28	4332.287	-2.55	0.011	-19549.88	-2560.675
b13_1	26090.95	28632.83	0.91	0.362	-30051.34	82233.25
b14_1	24822.03	24635.69	1.01	0.314	-23482.79	73126.85
b16_1	162.4914	144.5716	1.12	0.261	-120.9796	445.9625
_cons	2.36e+07	2864441	8.25	0.000	1.80e+07	2.93e+07

Petrol and Parking fees and tolls

. regress itm828 num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1						
Source	SS	df	MS	Number of obs = 19,464		
Model	2.8130e+12	10	2.8130e+11	F(10, 19453) = 383.89		
Residual	1.4255e+13	19,453	732783100	Prob > F = 0.0000		
				R-squared = 0.1648		
				Adj R-squared = 0.1644		
Total	1.7068e+13	19,463	876939642	Root MSE = 27070		
itm828	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	-163.4554	74.23394	-2.20	0.028	-308.9603	-17.95052
itml90	.0133648	.0003148	42.45	0.000	.0127477	.0139819
year	974.7846	58.80386	16.58	0.000	859.524	1090.045
b3_1	1155.815	488.9672	2.36	0.018	197.3978	2114.233
b4_1	-68.56681	20.75493	-3.30	0.001	-109.2483	-27.88537
b5_1	1205.589	119.2455	10.11	0.000	971.8572	1439.32
b11_1	-3322.191	178.8588	-18.57	0.000	-3672.77	-2971.613
b13_1	7834.824	1190.641	6.58	0.000	5501.065	10168.58
b14_1	5196.718	986.5496	5.27	0.000	3262.996	7130.44
b16_1	-48.30363	5.949578	-8.12	0.000	-59.96531	-36.64194
_cons	-1927220	117341.6	-16.42	0.000	-2157220	-1697221

Rent and Water expenditure

. regress itm760 num_stations itml90 year b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1						
Source	SS	df	MS	Number of obs = 29,850		
Model	9.2196e+13	10	9.2196e+12	F(10, 29839) = 816.50		
Residual	3.3693e+14	29,839	1.1292e+10	Prob > F = 0.0000		
				R-squared = 0.2148		
				Adj R-squared = 0.2146		
Total	4.2913e+14	29,849	1.4377e+10	Root MSE = 1.1e+05		
itm760	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	2796.582	263.0888	10.63	0.000	2280.916	3312.247
itml90	.0535792	.0010439	51.33	0.000	.0515331	.0556253
year	3598.093	167.2335	21.52	0.000	3270.308	3925.878
b3_1	11644.82	1574.05	7.40	0.000	8559.611	14730.03
b4_1	908.0323	62.98752	14.42	0.000	784.574	1031.491
b5_1	7100.308	364.5999	19.47	0.000	6385.676	7814.94
b11_1	-15202.06	589.8441	-25.77	0.000	-16358.18	-14045.94
b13_1	40717.13	3648.318	11.16	0.000	33566.27	47867.99
b14_1	17468.89	2871.149	6.08	0.000	11841.31	23096.46
b16_1	-150.064	18.7565	-8.00	0.000	-186.8276	-113.3005
_cons	-7073976	333388.4	-21.22	0.000	-7727432	-6420521

Water charge

. regress itm765 num_stations itml90 year area b3_1 b4_1 b5_1 b11_1 b13_1 b14_1 b16_1						
Source	SS	df	MS	Number of obs = 29,660		
Model	2.7606e+10	11	2.5097e+09	F(11, 29648) = 352.25		
Residual	2.1123e+11	29,648	7124681.67	Prob > F = 0.0000		
				R-squared = 0.1156		
				Adj R-squared = 0.1153		
Total	2.3884e+11	29,659	8052829.23	Root MSE = 2669.2		
itm765	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
num_stations	20.55418	6.622482	3.10	0.002	7.573821	33.53453
itml90	.00083	.0000263	31.58	0.000	.0007785	.0008815
year	136.4895	4.217663	32.36	0.000	128.2227	144.7563
area	.0899345	4.364142	0.02	0.984	-8.463976	8.643845
b3_1	-89.0607	39.72147	-2.24	0.025	-166.9165	-11.20487
b4_1	3.646307	1.594119	2.29	0.022	.5217632	6.770851
b5_1	-100.388	9.287452	-10.81	0.000	-118.5918	-82.18415
b11_1	-195.5313	14.91261	-13.11	0.000	-224.7606	-166.3019
b13_1	250.173	92.08389	2.72	0.007	69.68449	430.6615
b14_1	654.8965	72.44419	9.04	0.000	512.9027	796.8903
b16_1	-5.099749	.4731649	-10.78	0.000	-6.027173	-4.172324
cons	-268308.3	8400.743	-31.94	0.000	-284774.1	-251842.4

Questions

- Construction of DiD estimators
 - Year of treat 1 * treat1
 - Year of treat 2 * treat2
 - etc?
- How might I add the information of number of stations per district? Could see diminishing returns on specific outcomes as stations increase?
- Other districts in cities as control groups?