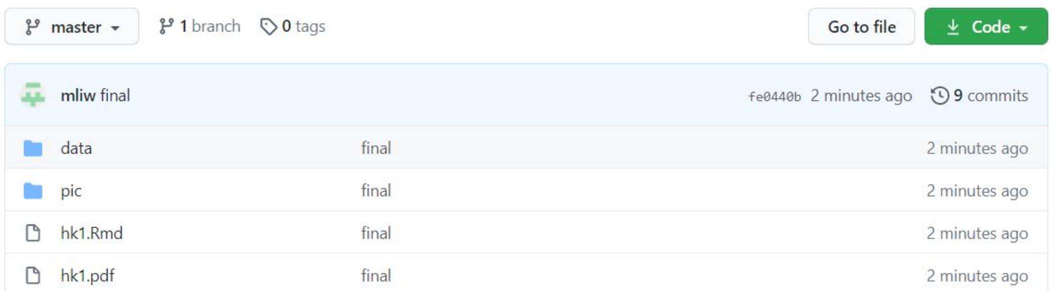
ECO395M STAT LEARNING Homework 3[*]

Mingwei Li, Xinyu Leng, Hongjin Long

Abstract

This document is the third homework of ECO395M STAT LEARNING.



[*]Mingwei Li, Xinyu Leng and Hongjin Long are master students of economics, The University of Texas at Austin

Contents

1 What causes what?

(Q1_1) Why can't I just get data from a few different cities and run the regression of "Crime" on "Police" to understand how more cops in the streets affect crime? ("Crime" refers to some measure of crime rate and "Police" measures the number of cops in a city.)

What we want to know is the causal effect of police on crime rate. This is of great policy significance.

It's possible that places with an inordinate amount of crime tend to employ a large police force. Which means crime rate is the causal effect of police.

The result of such regression can not credibly identify a causal effect of police on crime.

(Q1_2) How were the researchers from UPenn able to isolate this effect? Briefly describe their approach and discuss their result in the "Table 2" below, from the researchers' paper.

How were the researchers from UPenn able to isolate this effect? Briefly describe their approach

They use the easily identifiable and clearly exogenous shock provided by changes in the terror alert level in Washington, D.C., to evaluate the causal effect of police on crime. A notable benefit of their research design is that their treatment, the terror alert level, turns on and off repeatedly during their sample.

The logi is: changes in the terror alert level $\Rightarrow$ changes in the number of police $\Rightarrow$ changes in different types of crimes. Therefore, they can estimate the causal effect of police on crime.

and discuss their result in the "Table 2" below, from the researchers' paper.

The table 2 is listed below. We would discuss the entry in the table one by one.

TABLE 2

TOTAL DAILY CRIME DECREASES ON HIGH-ALERT DAYS

	(1)	(2)
High Alert	−7.316*	−6.046*
	(2.877)	(2.537)
Log(midday ridership)		17.341**
		(5.309)
R^2	.14	.17

NOTE.—The dependent variable is the daily total number of crimes (aggregated over type of crime and district where the crime was committed) in Washington, D.C., during the period March 12, 2002–July 30, 2003. Both regressions contain day-of-the-week fixed effects. The number of observations is 506. Robust standard errors are in parentheses.

* Significantly different from zero at the 5 percent level.

** Significantly different from zero at the 1 percent level.

(row1 column1 -7.316): The results from their most basic regression are presented in Table 2, where they regress daily D.C. crime totals against the terror alert level (1 high, 0 elevated) and a day-of-the-week indicator. The coefficient on the alert level is statistically significant at the 5 percent level and indicates that on high-alert days, total crimes decrease by an average of seven crimes per day, or approximately 6.6 percent

(column2 -6.046 17.341): To investigate the effect of tourism more systematically, in column 2 of Table 2 they verify that high-alert levels are not being confounded with tourism levels by including logged midday Metro ridership directly in the regression. The coefficient on the alert level is slightly smaller, at -6.2 crimes per day. Interestingly, they find that increased Metro ridership is correlated with an increase in crime. The increase, however, is very small—a 10 percent increase in Metro ridership increases the number of crimes by only 1.7 per day on average. Thus, given that midday Metro ridership is a good proxy for tourism, changes in the number of tourists cannot explain the systematic change in crime that they estimate.

(Q1_3) Why did they have to control for Metro ridership? What was that trying to capture?

Why did they have to control for Metro ridership?

What has been confirmed from basic regression is high alert level ⇒ less crime.

There are 2 hypotheses:(1) high alert level ⇒ high police level ⇒ less crime. (2) high alert level ⇒ less tourism ⇒ less crime. They want to rule out the second hypothesis, therefore they have to control for Metro ridership.

What was that trying to capture?

They are trying to capture the causal effect of tourism on crime.

(Q1_4) Below I am showing you "Table 4" from the researchers' paper. Just focus on the first column of the table. Can you describe the model being estimated here? What is the conclusion?

TABLE 4

REDUCTION IN CRIME ON HIGH-ALERT DAYS: CONCENTRATION ON THE NATIONAL MALL

	Coefficient (Robust)	Coefficient (HAC)	Coefficient (Clustered by Alert Status and Week)
High Alert × District 1	−2.621**	−2.621*	−2.621*
	(.044)	(1.19)	(1.225)
High Alert × Other Districts	−.571	−.571	−.571
	(.455)	(.366)	(.364)
Log(midday ridership)	2.477*	2.477**	2.477**
	(.364)	(.522)	(.527)
Constant	−11.058**	−11.058	−11.058+
	(4.211)	(5.87)	(5.923)

NOTE.—The dependent variable is daily crime totals by district. Standard errors (in parentheses) are clustered by district. All regressions contain day-of-the-week fixed effects and district fixed effects. The number of observations is 3,542. R^2 = .28. HAC = heteroskedastic autocorrelation consistent.

+ Significantly different from zero at the 10 percent level.

* Significantly different from zero at the 5 percent level.

** Significantly different from zero at the 1 percent level.

Can you describe the model being estimated here?

D.C has many districts. District 1 is the most important one, as White House is there. Therefore, the police would place a great amount of force in district 1 during high-alert period.

The regression with district fixed effects is in Table 4. During periods of high alert, crime in the National Mall area(district 1) decreases by 2.62 crimes per day. Crime also decreases in the other districts, by .571 crimes per day, but this effect is not statistically significant.

What is the conclusion?

Police has a negative causal effect on crime, after controling other factors similar across the districts.

We assume the police level in district 1 is much higher than other districts. In this case, the difference between the High Alert×District One and the High Alert×Other Districts coefficients is a differencein-difference estimator that controls for all common factors between the districts. If bad weather, for example, causes decreases in crime, a coincidental correlation with the timing of a high alert could confound their results. The difference-in-difference estimator controls for any factors such as weather, tourism, or other events that affect the districts similarly. Even after controlling for all such factors and recognizing that their assumption is too strong, they still find that crime decreases in District 1 during high-alert periods by some two crimes per day, or more than 12 percent.