Economics and Crime

Finding Statistical Inference

Michael Whittlesey

Table of Contents

Abstract:

The purpose of this report is to determine what possible causes, if any, may account for increasing crime trends. These possible causes have been isolated to only include socioeconomic variables that are deemed reliable in their estimation. This research will be done only for the city of Seattle WA; however, it will be structured in order that the process can be repeated in other cities in order to aid in subsequent quasi-analysis. Given that a socioeconomic variable is found to have a significant correlation to increasing crime trends, I will argue for or against the validity of a causal relationship in lieu of a correlation. My research plan includes the following stages: 1) researching theories of notable experts in the field, 2) empirical analysis of raw data and analysis of quantified results, 3) interview of law enforcement official along with researching of historical validity, 4) recommendation of further study and actions the public can take to prevent increasing crime trends. The first stage will encompass finding a previously established model that I can base my research off of. The second will include finding data in an array format which I can then reformat into a time series. These time series or variables will then be applied to a multiple regression model after adequate transformation of the variables (logarithmic, auto regression, or seasonal decomposition). After I have created an adequate model the results will be analyzed and tested using a simple F-Test. Given the results of the model, I will interview a select law enforcement official with experience in order to provide some clarity as to the validity of a causal relationship between the economic variables and their effect on increasing crime trends. The same process will be repeated by looking back through history to find occurrences that either prove or disprove my findings. Lastly, I will provide recommendations for future analysts in repeating the experiment in other cities, and for the public in what will prevent increasing crime trends.

Introduction

Crime has always been a problem for society. Over the course of human history there have been plenty of instances in which the level of criminal activity rose to an uncontrollable level. One example of this calamitous process of increasing social upheaval was observed in the events leading up to the French Revolution. Late nineteenth century England provides another equally applicable illustration (Staff, 2009). Clearly, it would be prudent to find the cause of these surges in social lawlessness. Subsequently, this purpose led to my subject of finding the answer to rising criminal activity in society. In order to simplify the research process I isolated my analysis to only include variables applied to the city of Seattle WA. The following analytical report details my struggle to answer the aforementioned question: *What socioeconomic factors correlate to increasing levels of crime? And subsequently, is it possible to reduce criminal activity by modifying these socioeconomic factors.*

Many of us would prefer to ignore the darker corners of society. Despite its' unpleasant aftertaste, it is important to take a focused and unbiased look at crime and its' causes. And, if we do not learn from our mistakes, i.e. the anarchism of the French Revolution, we are doomed to repeat history and cause even more human suffering and injustice. The main point of this report is to reveal the controllable variables that lead to a lawless society, so that they may be avoided. Avoiding the plunge into a lawless society is something rarely stressed in society; however, it is of great importance. This lack of adequate attention is due to the fact that the possibility of rampant crime is easy to ignore when it doesn't pose a threat on frequent basis at the personal level. However, from personal experience (see appendix F) it is present in small subsections of our society, lying closer than one would assume.

This question concerning the cause of increased lawlessness has been the focus of many researchers astronomically more notable than myself. In fact, Economist Gary Becker won the Nobel Prize in 1992 for his work in analyzing the incentives driving criminal behavior (Becker, 1992). His most notable accomplishment was coined the "Economic Model of Crime". A model which estimated an individual's time partaking in criminal activity based on multiple contributing factors. It is convenient and perhaps even comforting to think of criminals acting in such a rational fashion. However, it is often the case that economists assume that the rest of society thinks as they do. In turn, these social scientists concoct theories centering around the premise that people on average make decisions after a careful contemplation of the immediate cost, opportunity cost, and possible benefits. Gary Becker attempted to validate his model given the resources available, however it is still widely accepted as only a theory. Furthermore, his model applies only to distinct observation (individual criminals).

The following report consists to of four sections. These sections are the following:
1) methodology, 2) results, 3) discussion of my results and 4) conclusions and recommendations.

Methodology

For the sake of simplicity, I narrowed my focus to include only the city Seattle WA, the data was obtained through (data.seattle.gov, 2016). I took the approach of comparing observed criminal activity to the city's economic state. If Becker's theory remains valid, there should be a notable relationship between economic indicators and criminal activity. Throughout my research, I considered a wide range economic factors. Examples of such economic indicators include the following: unemployment, GDP per capita, income, and the consumer price index (CPI). My strategy assumes that for any given society the level of criminal activity is dependent on the economic well-being of that society.

I have presented my research in the following order in order that my findings are clear and comprehensible. First, I will detail the methodology leading me to choose Becker's Economic Model of Crime as an outline for my own model. Second, I will provide the results I obtained from analyzing Seattle's crime and economic data. The third section will discuss the implications of my results along with how they align with historical data and personal accounts of law enforcement officers. After which, I will conclude with recommendations for the future.

Section 1: Background Research on Crime Behavioral Models

In Becker's award winning publication "Crime and Punishment: An Economic Approach", he details his economic model of crime (Becker, 1974). His main argument states that despite there being a huge array of criminal activities, there are common properties behind the incentives for committing such illegalities. Becker was attempting to bring light to what the optimal level of punishment concerning the enforcement of laws against actions deemed as criminal behavior. According to Becker, In order to find this optimal level, one must first find a means of measuring the cost crime, given its level of activity. The first model Becker provides concerns the explicit and implicit cost of the harm done by the illicit action. The following math may seem dry and redundant, but understanding the base of this crime analysis theory is the crucial reason why my methodology and succeeding model has validity and importance.

$$H_i = H_i(O_i)$$

The adjacent model is basically stating that the cost of harm 'Hi' is directly related to the activity level 'Oi' of that illicit activity.

This leads to the second major model formed by Becker, which incidentally, is the cornerstone of the final crime model.

$$G_i = G_i(O_i)$$

Essentially, the adjacent model states that just as increased criminal activity gleans a higher social cost it also increases the personal gain of the offender.

Finally we come to the culmination of these two models which shows the net cost of a criminal act at variable levels of activity. Ultimately this net cost to society is the difference between the harm stemming from the illegal action and the gain of the offender (remembering that the offender is also part of society).

$$D(O) = H(O) - G(O)$$

The adjacent model simply shows the difference of the above two models.

This model is often used as a means of finding the optimal level of enforcement cost to be set by public policy. However, it hinges on one key factor, that each of these crime is committed with the personal gain of the offender in mind.

The aforementioned Economic Model of Crime, is based off of the previous equations. Moreover, it takes into account the cost of committing the crime in the form of risk along with the opportunity cost of not committing the crime. The major difference between this algorithm and my own is that Becker's model considers each observation to be single individual. Whereas I built my model to measure the aggregate level of crime in relation to the opportunity cost of not engaging in criminal activity.

$$y = f(x_1, x_2, x_3, x_4, x_5, x_6, x_7)$$

Refer to below table for description of variables

As you can see from the adjacent chart *Figure 1*, one of the major variable effecting whether a person will or will not commit a crime is the wage they could be receiving from legal employment. Therefore, the measure of employment should have an effect on criminal activity, in an isolated area. This is why I have chosen to measure the level of criminal activity in the city of Seattle and plan to find its correlation to the level of unemployment and other related economic variables. If Becker was correct in his theory of what motivates a criminal to act, there should be some correlation between crime and level of economic health. Moreover, this correlation will be most notable in areas where financial gain was definitely a factor in the type of crime committed.

Variable	Description
y	hours spent in criminal activities
x_1	"wage" for an hour spent in criminal activity
x_2	hourly wage in legal employment
x_3	income other than from crime or employment
x_4	probability of getting caught
x_5	probability of being convicted if caught
x_6	expected sentence if convicted
x_7	age

Figure 1

Section 2: Analyzing Crime Data and Finding Correlating Variables

Though I prefer the simplistic term of 'data analysis', the academic term for any process in which statistical methods are applied to economic data is dubbed as "econometrics". However, this cumbersome word is hardly recognizable. The most simplistic and accurate description of this study is as follows: (E.Malinvaud, 1966) "The art of the econometrician consists in finding the set of assumptions that are both sufficiently specific and sufficiently realistic to allow him to take the best possible advantage of the data available to him." (p. 514). For other terms of a similar unrecognizable nature please see Appendix C: Glossary. Consequently, my empirical analysis begins with re-organizing a rather large data frame of more than 250,000 observations of historical crime data into their respective placement along a monthly time sequence. All of the data along with their online source can be found in Appendix E: Visuals of Key Variables. This time series format is visualized in *Figure 2* below. The jagged orange/blue line presents a measurement of the number of monthly reports of Robbery for the city of Seattle WA for years 2008 through the beginning of 2014. The orange segments represent months with aggregate crime that was significantly above average. Whereas the blue segments represent months with lower than average aggregate crime. The smoothed trend line was calculated using Loess Smoothing methods and shows a heightened effect of seasonally decomposing a time series. (Hyndman, 2016)

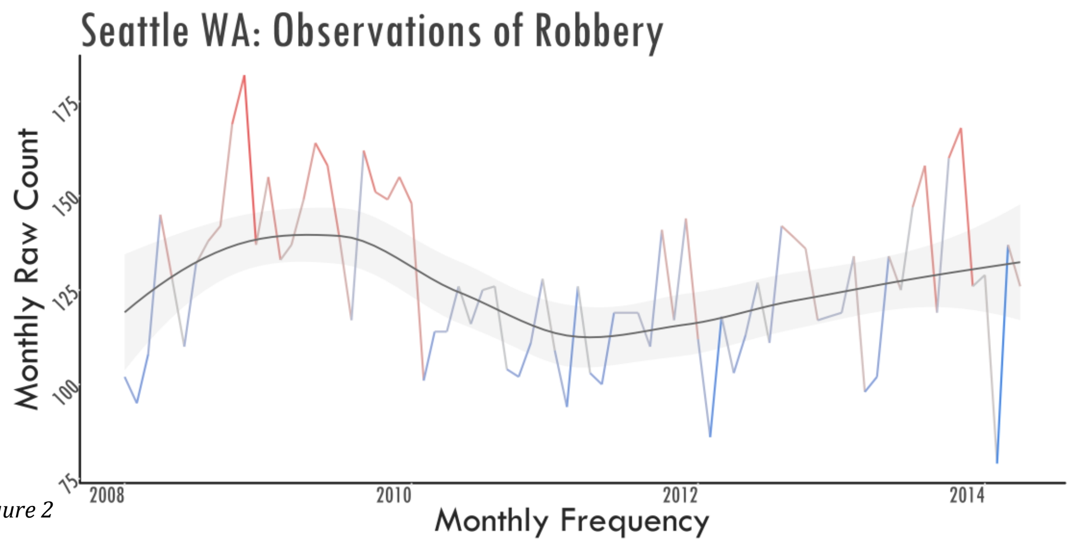


Figure 2

As you can quickly observe from the color scale, crime varies substantially within a relatively short amount of time. And if you compare the variation of crime with that of population below *Figure 3* (same color scale), it is obvious that the crime level is not randomly dependent on the number of people in a given location.

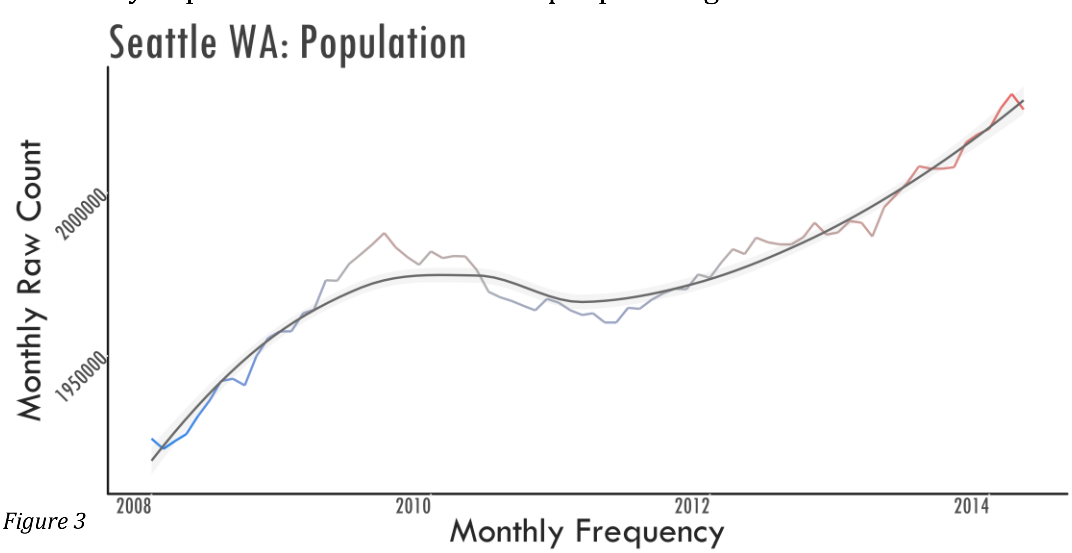


Figure 3

Due to the complicated nature of analyzing time series data, the variable of “time” must be accounted for. To compensate, I used STL techniques as stated in Forecasting Principles and Practices (Hyndman, 2016, p. 6.5). The stationary, time series variables were than applied to a logarithmic multiple regression model. This shows the percentage change in the dependent variable (crime) given a change in the independent variables (economic indicators). Logarithms are very useful in that they offer clear interpretation of results. For example, if a one percent increase in unemployment causes an increase of 200 more burglaries per month, that is non-descriptive. Such results don’t detail whether this is a significant increase or fairly minor one. This is especially important when dealing with variables in a time sequence. Throughout this process, my focus was on deciding which economic variables I was to use for my model. This is a intricate process, these explanatory variables must be completely independent of one another, or as close to that state as possible (Wooldridge, 2013, p. 345). Furthermore, if the results are to be accurate, all of the variables causing increases in crime must be accounted for and present in the equation. Obviously, it is not possible to account for everything, or even collect such data in many cases. This process of deciding which variables to use and which are to be omitted is highly scrutinized. For further information on issues relating to this topic and the debate among econometricians please see Appendix D: Ethical Consideration.

After a considerable amount of deliberation I selected the economic variables that were to be used in my model along with isolating the dependent variables of crime. The crime variables are as follows: Burglary, Robbery, and Assault. These could not simply be added together because of the overlap in the judicial system. The economic variables showing the most independence from one another along with a logical connection to crime were than chosen as the following: the consumer price level or CPI, unemployment as a percentage of the labor force, population, and an economic growth index acquired from the Federal Reserve Bank of St. Louis (FRED, 2016). After finalizing my model, the subsequent step was to interpret the results.

Note: All of my work concerning the reformatting, transforming, and analysis of the data can be seen in Appendix G: R-Markdown.

Results

The results of the aforementioned model were largely unexpected. My initial assumption was that unemployment would have the largest impact on increasing crime. However as you can see from the adjacent table *Figure 4*, increasing prices or CPI has the largest and most consistent effect overall. These results concerning CPI can be interpreted as the following:

A 1% increase in prices (CPI) causes a

- 3% increase in Assaults
- 4% increase in Burglaries
- 5% increase in Robberies

In comparison, a 1% increase unemployment causes less than 1% increase in all three crimes. However, it is notable that most of the estimated economic variables are in the logical direction (i.e. positive vs. negative). A curious, though not illogical, exception is the percentage increase in population. Both assault and burglaries show a decrease in crime with increasing population given that all other factors are held constant. Robbery on the other hand shows a significant increase in crime 6.5% given a 1% increase in population. These contradictions are important in that they point towards the need for further refinement of the model. This is the primary reason why these crime variables were isolated and chosen. Burglary can be categorized as low risk - high reward, Robbery as high risk - medium reward, and Assaults as low risk - no reward. Any consistency among the results provides a more accurate measure of a possible economic causal relationship to increasing crime, due to the variance of the crime variables. The primary way in which these models are evaluated is through mathematical tests. These tests measure the difference between what was accounted for and what occurred randomly (random from a mathematical perspective). The F-Statistic at the bottom of the table *Figure 4* details whether the model, as a whole, is considered significant. The value of the F-Statistic deems a models collective interaction among variables as either significant or insignificant. Next to this value on the table there are three asterisks denoting that each model is highly significant. This significance means that the model cannot be disproven, it does not establish proof (Wooldridge, 2013, p. 153)

Assault

	Dependent variable:
	as.numeric(log(sa.seattle_all[[n]]))
cpi)	2.924
	(2.230)
unemployment_lag	0.041***
	(0.014)
econ_growth_index	0.008
	(0.009)
pop)	-1.949
	(2.654)
count	0.001
	(0.004)
Constant	17.043
	(42.411)
Observations	76
R^2	0.405
Adjusted R^2	0.362
Residual Std. Error	0.095 (df = 70)
F Statistic	9.527*** (df = 5; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

Burglary

	Dependent variable:
	as.numeric(log(sa.seattle_all[[n]]))
cpi)	3.792**
	(1.823)
unemployment_lag	0.016
	(0.012)
econ_growth_index	-0.002
	(0.008)
pop)	-0.539
	(2.170)
count	-0.002
	(0.003)
Constant	-6.531
	(34.677)
Observations	76
R^2	0.238
Adjusted R^2	0.183
Residual Std. Error	0.078 (df = 70)
F Statistic	4.367*** (df = 5; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

Robbery

	Dependent variable:
	as.numeric(log(sa.seattle_all[[n]]))
cpi)	5.057**
	(2.431)
unemployment_lag	0.007
	(0.016)
econ_growth_index	-0.022**
	(0.010)
pop)	6.591**
	(2.894)
count	-0.008**
	(0.004)
Constant	-117.942**
	(46.242)
Observations	76
R^2	0.483
Adjusted R^2	0.446
Residual Std. Error	0.104 (df = 70)
F Statistic	13.079*** (df = 5; 70)
Note:	*p<0.1; **p<0.05; ***p<0.01

Figure 4

Overall, increasing price is the most notable and consistent of the economic variables chosen for my model. Therefore, the remainder of this report will focus on whether evidence can be found to support or disprove my results, and how we can use this theory to prevent future surges in crime.

Discussion

We have established that percentage increase in prices or log(CPI) has a positive statistically causal relationship with increasing crime levels. These results were originally puzzling simply because I have never heard the two variables equated in such a relationship. In fact the overall consensus is that unemployment has a larger effect on crime than increased price level (Gupta, 2016). And if you mull over the implications of increased prices as the culprit rather than unemployment, you quickly realize that this is a very "politically incorrect" idea. CPI and specifically percentage change in CPI is often called inflation. And if one of the key sources of inflation is government spending, regulation and taxation, it is easy to see where that road leads to. However, in this report I refer to percentage increase in CPI as simply "increase in prices". This is due to the fact that inflation is normally associated with government sanctioned increases of the money supply due to over spending. I am holding that there is a distinction between the two phenomena and proving a relation is another research project entirely.

Given the potential unpopularity of my results, my first step was to consider if my math could be backed up by historical evidence. Turns out, history has a few such examples. During the late 18th century, France was in a state of turmoil and on the cusp of a revolution. However, it was government regulation of the bread supply that acted as the catalyst which plunged the country into a state of mass lawlessness and bloodshed (Staff, 2009). This supply regulation caused the price of bread to increase to a high enough level that the public simply resorted to violence and force to obtain what they needed. Another, equally notable example, took place in 1921 Germany. Germany accrued a large amount of debt financing WWI, and when England demanded payment, the German mark lost value at an exponential rate. This phenomena is referred to as hyper-inflation. Historical records from that time state that workers were paid multiple times a day in order that they could spend the money before the currency lost further value. Moreover, crime rose to a rampant rate as the public resorted to thievery and prostitution in order to obtain essential items (Litle, 2013). These are of course extreme examples. However, upon interviewing a law enforcement agent with a wealth of experience I was briefed on yet another instance where my theory appears to remain valid. For a full transcript of the interview see Appendix B: Interview with Gene Martin (Martin, 2016). Gene Martin was a member of WA State Patrol during the 1970s oil crisis. Throughout this crisis gasoline prices rose so drastically that there was a surge in the crime of gas siphoning. The crime became so rampant that law enforcement was forced to put locks on the gas caps of their patrol vehicles. Moreover, I am not the first to come to this conclusion via statistical inference. A study done in the United Kingdom by Chor Foon Tank and Hooi Hooi Lean is quite similar to my own in their methodology and results (Tang & Lean, 2007). However, I was only able to obtain the abstract due to the cost of obtaining the entire publication.

Overall, the above evidence is not enough to come to a clear conclusion. The possibility of making a mistake in an econometric study far outweighs the likelihood of getting it right on the first try. However, there are still many avenues of approach for further progress towards proving a causal relationship between crime and increased prices.

Conclusion

In conclusion the main purpose of this research report was to find a possible causal link between crime and economic variables. This study was done in the hopes of finding an economic variable that would allow prevention of sudden increases in crime. And a possible causal link has been identified as increased prices. Moreover, there are proactive steps society can take, given that my model and following results are accurate.

Recommendations

Basic economic theory claims a plethora of reasons for sudden increases in prices. For the most part, these sudden increases are due to either a decrease in the supply of a product, or an increase in the demand for that product. In order to prevent these sudden shifts in demand and supply it is important to enact policies that will cause such shifts at a slow, marginal rate. Take for example, the soon to be increase in minimum wage. Few economists will argue against the notion that given an increase in minimum wage, prices will increase as well. This is simply a result of people having more money and those selling products being able to charge a higher price. The current plan is a steep increase in the minimum wage from $9.47 to $11.50. At first everything will be fine, people will have more money and shop owners will be under the impression that there is a higher demand for their product. However, given a time lag, prices will rapidly increase as shop owners compete with one another to increase prices and gain a higher profit to compensate their own losses of labor costs. I suggest that we enact this at an even slower rate, even if it comes down to 5 cents per month, that will smooth out any sudden increase in price. Furthermore, it will give the much needed aid to those who require it. And from the historical evidence provided it is only the sudden increases that pose a problem of increasing crime levels. This same argument can also be applied to taxes or monopolistic firms exacting more revenue than they otherwise should. For example, pharmaceutical companies have been steadily increasing prices to take advantage of the new affordable health care act. According to my theory, increasing the price of a necessity such as medical care could have a drastic effect on crime.

On a final note, in order to improve my model's validity and real-world use, further steps must be taken. The first of which is to repeat the process throughout similar cities in the United States and observing whether the results are consistent. Crime may not be the most profitable of data science genres, but it is important. It is crucial that we find an answer to preventing these surges in crime; in order for society to recognize the problem and its' solution before it is too late.

Appendix A: Work Cited

Becker, G. S. (1974). *Volume Title: Essays in the Economics of Crime and Punishment.* NBER.(Print, Background)

data.seattle.gov. (2016, December 07). Retrieved from Seattle | Open Data: https://data.seattle.gov/(Electronic/Raw Data, Exhibit)

E.Malinvaud. (1966). *Statistical Methods of Econometrics.* Chicago: Rand McNally & Co.(Print, Method)

FRED. (2016, December 07). *Economic Data*. Retrieved from Economic Research Federal Reserve Bank of St. Louis: https://fred.stlouisfed.org/(Electronic/Raw Data, Exhibit)

Gujarti, D. (2003). *Basic Econometrics, International Edition.* NA: McGraw-Hill Higher Education.(Print, Method)

Gupta, P. (2016). How unemployment affects serious property crime: A national case-control study. *Journalist Resource*.(Electonic, Argument)

Hyndman, R. J. (2016). *Forecasting: principles and practice.* OTexts.(Print, Method)

Litle, J. (2013, August 25). *An Important History Lesson: Hyperinflation in the Weimar Republic.* Retrieved from Economic Policy Journal: http://www.economicpolicyjournal.com/2013/08/an-important-history-lesson.html(Electronic, Argument)

Martin, G. (2016, November 20). Interview. (M. Whittlesey, Interviewer)(Print/Raw Data, Exhibit/Argument)

Staff, H. (2009). *History.com*. Retrieved from History.com: http://www.history.com/topics/french-revolution(Electronic, Argument/Background)

Tang, C. F., & Lean, H. H. (2007, November). *National Criminal Justice Reference Service*. Retrieved from NCJRS: https://www.ncjrs.gov/App/Publications/abstract.aspx?ID=243186(Electronic/Background

The Economist. (2005, 11 1). Retrieved from The Economist: http://www.economist.com/node/5246700(Electronic, Background)

Wooldridge, J. M. (2013). *Introductory Econementrics.* Mason, OH: Cengage Learning.(Print, Method)

Appendix B: Interview Transcript for Gene Martin

November 20, 2016. Gene Martin was given overview of project prior to interview

Interviewer (MW): Can you give me an overview of your experience in law enforcement?

Gene Martin (GM): Well, I was a Washington state patrol officer for 36 years. And after that I worked in the Department of Transportation dealing in motor vehicle theft. Most of my time in law enforcement we spent doing traffic control. However, as State Patrol I was privy to the information of all other cases going on in the state and was able to assist if need be.

Interviewer (MW): Excluding domestic violence, rape and murder, would you categorize most crimes as being financially motivated?

Gene Martin (GM): I would definitely say that the majority of crimes have a financial motivation whether it is to get money for drugs or simply because they want something that someone else has. Take burglary for example, often those people know the people that they are robbing from. And they simply look for the opportunity to take something that they wanted.

Interviewer (MW): Would you say that being unemployed is a common factor among criminal?

Gene Martin (GM): Definitely, more often than not the people we arrested were unemployed. However, I would not necessarily say that they were recently unemployed. Especially since most of the people arrested are previous criminals, they lived most of their lives after that point without a job. And they use crime to supplement any other income they might have.

Interviewer (MW): My results showed a high correlation between increasing prices and increasing crime. Especially with robbery and burglary. This puzzles me, do you have an explanation?

Gene Martin (GM): Yes I actually have a perfect example. During the 1970s oil crisis gas, prices went up super high and people resorted to siphoning gas, we ended up having to put locks on are gas caps for patrol vehicles because it got so bad. And if we go back to burglary it would make sense that people would be more likely to simply take something rather than pay for it if the price was just too high. In that respect your results make quite a bit of sense.

Interviewer (MW): Why would motor vehicle thefts not show this same trend?

Gene Martin (GM): Motor vehicle theft can be a bit confusing if you're trying to find out what's motivating it. Sometimes it's not entirely motivated by money. Lot of the times cars are stolen with the intention of committing other crimes like drive-by shootings. These cars are then dumped, if someone wants to sell one of these cars they are going to need to strip it for parts or get rid of anything that can be traced back to them or that the car is stolen.

Appendix C: Glossary

Economic Variables: A measurement that determines how an economy functions. Examples include population, poverty rate, inflation, and available resources.

F-Test: The hypothesis that the means of a given set of normally distributed populations, all having the same standard deviation, are equal. This is perhaps the best-known F-test, and plays an important role in the analysis of variance (ANOVA).

Logarithmic Transformation: The process of taking a logarithmic value of a vector of numbers, resulting in the percentage increase.

Null Hypothesis: The hypothesis that there is no significant difference between specified populations, any observed difference being due to sampling or experimental error.

P-Value: The probability of obtaining a result equal to or "more extreme" than what was actually observed, when the null hypothesis is true.

Quasi-Experiments: Experiments designed to compensate for internal bias. In the case of this study it entails the comparison of the results of a singular time series study and the application to parallel populations.

Regression Analysis: A statistical process for estimating the relationship among variables. If there are more than one variables the regression becomes “multiple”.

Seasonal Decomposition: a statistical method that deconstructs a time series into several components, each representing one of the underlying categories of patterns.

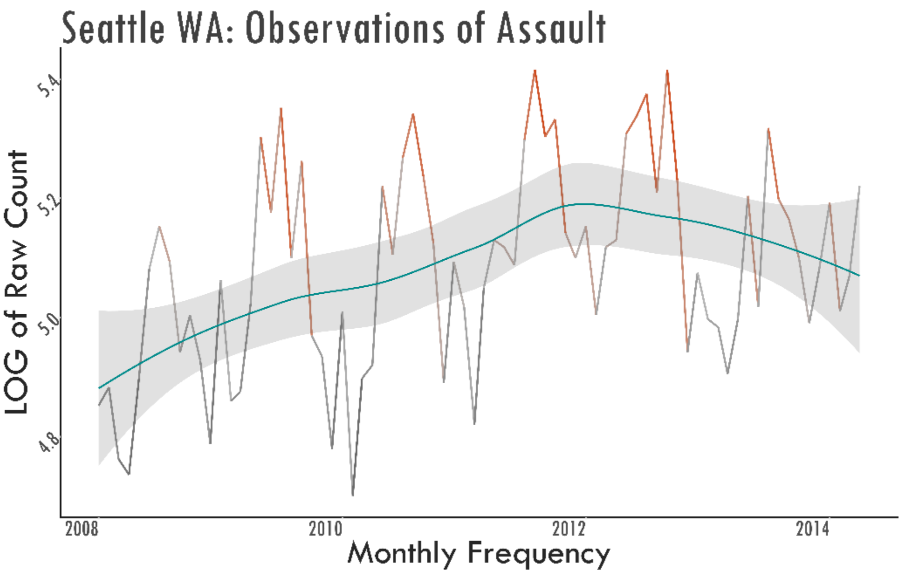
Time Series Trend: A trend exists when there is a long-term increase or decrease in the data. It does not have to be linear. Sometimes we will refer to a trend “changing direction” when it might go from an increasing trend to a decreasing trend.

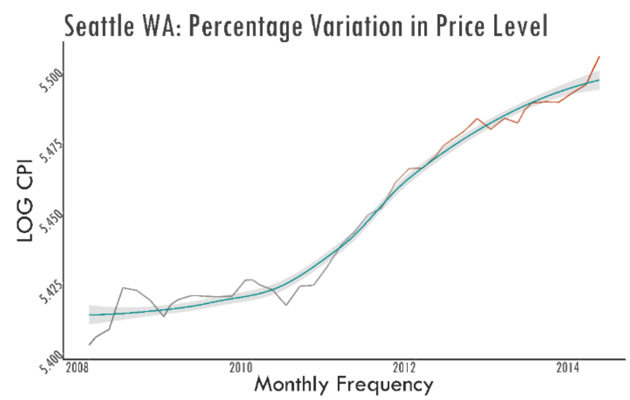
Appendix D: Ethical Consideration

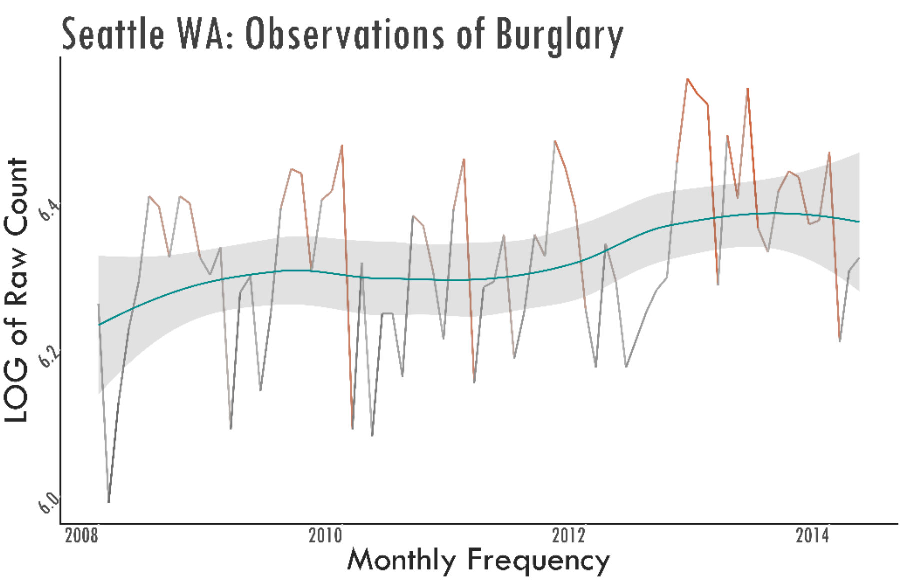
Before beginning, I would like to note some ethical considerations that come into play when inferring causality. Often, studies such as these make assumptions that are false and bias. This is especially true when attempting cross sectional time series analysis. Nothing is certain, and my results have a high possibility of being complete rubbish. In metrics the objective is to establish causality based on statistical inference. When applied to human behavior, the results are never 100% certain. And, there have been such examples.

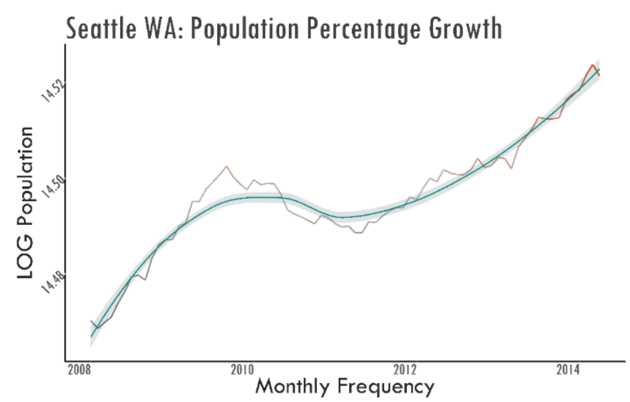
The author of best seller "Freakonomics ", Stephen J. Dubner, is a primary example of this situation. He published a report stating that there was a causal relationship between increased abortion and decreased violent crime. However, after two other economists from the Federal Reserve Bank of Boston attempted to repeat the experiment, they found a crucial error in Dubner's computer code. Subsequently, the causal relationship of abortion and crime fell apart. (The Economist, 2005)

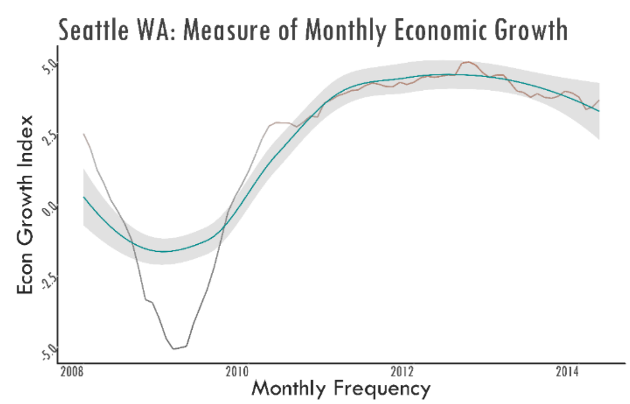
Appendix E: Visuals of Key Variables

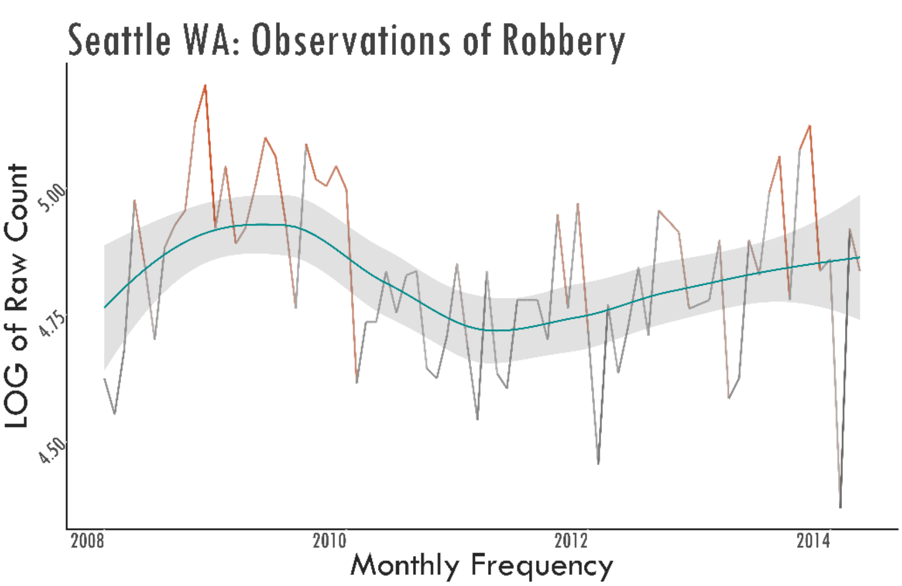


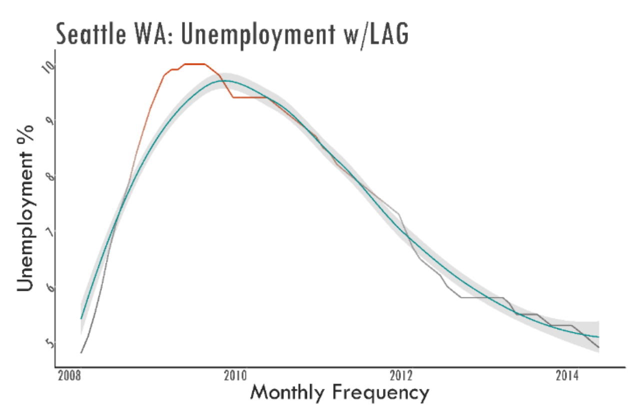












Appendix F: R-Markdown

The following shows only the code crucial to repeating my end result as well as creating some of the visuals throughout my report. If you wish to repeat this process, understand that the breakpoints within the code each include a user-created function/program which will allow for the most efficient memory allocation and possible modification. These functions are annotated as follows: "ez." & "<function name>". All aesthetics were processed through the package "Cairo" in order to increase dpi.

Seattle_Main_Progress_01.R

MIKE

Tue Dec 06 10:35:36 2016

```
#]---------------------------------------------------------[+++]
#]-----[ LIBRARY_IMPORT ]-----
library(readr)
library(plyr)
library(ggplot2)
library(zoo)
library(extrafont)
library(extrafontdb)
library(RColorBrewer)
library(fma)
library(stargazer)
#]-----[ DATA_IMPORT_MAIN ]-----
Seattle_Crime_2008_2014 <- read_csv("E:/R_DIRECT_01/R_01_DATA/Seattle_Crime_2
008-2014.csv")

## Parsed with column specification:
## cols(
##   `Police Beat` = col_character(),
##   CRIME_TYPE = col_character(),
##   CRIME_DESCRIPTION = col_character(),
##   STAT_VALUE = col_integer(),
##   REPORT_DATE = col_character(),
##   Sector = col_character(),
##   Precinct = col_character(),
##   Row_Value_ID = col_integer()
## )

#]-----[ REFORMAT_DATE ]-----
ez.allday_01 <- function(date_vec){
  dfrm_new <- data.frame(date_vec)
  dfrm_new$date <- as.Date(date_vec, "%m/%d/%Y")
  dfrm_new$month_num <- format(dfrm_new$date, "%m")
```

```
  dfrm_new$day_edit <- "01"
    dfrm_new$year <- format(dfrm_new$date, "%Y")
    dfrm_new$date <- as.Date(ISOdate(format(dfrm_new$date, "%Y"),format(dfrm_
new$date, "%m"),"01"))
    dfrm_new$date
  }
#]----->
Seattle_Crime_2008_2014$date <- ez.allday_01(Seattle_Crime_2008_2014$REPORT_D
ATE)
#]-----[ DATA_FRAME -> LIST_TIME_SEQUENCE ]-----
ez.df_list <- function(dfrm){

    lst <- as.character(count(dfrm[[1]])[[1]])

    new <- list()

    for(i in 1:length(lst))
      new[[i]] <- aggregate(
        subset(dfrm, dfrm[[1]]==lst[i])[[2]],
        list(subset(dfrm, dfrm[[1]]==lst[i])[[3]]),
        sum)

    names(new) <- c(lst[1:length(lst)])

    new
}
#]----->
seattle_all <- ez.df_list(Seattle_Crime_2008_2014[c(2,4,9)])
#]-----[ PLOTTING FUNCTIONS FOR VISUAL METRICS ]-----
ez.splot <- function(seattle_all, n, SMOOTH=TRUE, col1="grey15"){

  lst <- names(seattle_all)

  s1 <- data.frame(seattle_all[n])

  names(s1) <- c("Group.1", "x")

  ttl <- paste("Seattle WA: Observations of", lst[n], sep=" ")

  gtheme <- theme(
    plot.title = element_text(size=50, family="Tw Cen MT Condensed", colour="
gray25"),
    axis.text.y = element_text(angle=45, hjust=1, vjust=.4, size=30, colour="
gray30", family="Tw Cen MT Condensed"),
    axis.text.x = element_text(angle=0, hjust=1, vjust=0, size=30, colour="gr
ay30", family="Tw Cen MT Condensed"),
    axis.title.x = element_text(size=45, family=  "Tw Cen MT" , colour = "gra
y15"),
    axis.title.y = element_text(size=45, family=  "Tw Cen MT" , colour = "gra
y15"),
```

```
    axis.ticks = element_line(colour="gray90", size=.8),
    axis.line.x = element_line(color="gray10", size=1),
    axis.line.y = element_line(color="gray10", size=1),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_rect(fill="white"),
    legend.background = element_blank(),
    legend.title = element_text(size=30, face="bold.italic", colour="black"),
    legend.position = c(.10,.85))

  g1 <-
    ggplot(s1, aes(x=s1$Group.1, y=s1$x)) + geom_line(aes(x=s1$Group.1, y=s1$
x), colour=col1, size=.9) +
    xlab("Monthly Frequency") +
    ylab("Raw Count") +
    ggtitle(ttl) +
    geom_smooth(colour="darkcyan", alpha=.2, size=.5)+
    gtheme

  if(SMOOTH==FALSE){
    g1 <- ggplot(s1, aes(x=s1$Group.1, y=s1$x)) + geom_line(aes(x=s1$Group.1,
y=s1$x), colour=col1, size=.9) +
      xlab("Monthly Frequency") +
      ylab("Raw Count") +
      ggtitle(ttl) +
      gtheme
  }

  g1

}
ez.splot_log <- function(seattle_all, n, SMOOTH=TRUE, col1="dimgrey", col2="o
rangered3"){

  lst <- names(seattle_all)

  s1 <- data.frame(seattle_all[n])

  names(s1) <- c("Group.1", "x")

  ttl <- paste("Seattle WA: Observations of", lst[n], sep=" ")

  gtheme <- theme(
    plot.title = element_text(size=65, family="Tw Cen MT Condensed", colour="
gray25"),
    axis.text.y = element_text(angle=45, hjust=1, vjust=.4, size=30, colour="
gray30", family="Tw Cen MT Condensed"),
    axis.text.x = element_text(angle=0, hjust=1, vjust=0, size=30, colour="gr
ay30", family="Tw Cen MT Condensed"),
    axis.title.x = element_text(size=45, family=  "Tw Cen MT" , colour = "gra
```

```
y15"),
    axis.title.y = element_text(size=45, family=  "Tw Cen MT" , colour = "gra
y15"),
    axis.ticks = element_line(colour="gray90", size=.8),
    axis.line.x = element_line(color="gray10", size=1),
    axis.line.y = element_line(color="gray10", size=1),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    panel.background = element_rect(fill="white"),
    legend.background = element_blank(),
    legend.title = element_text(size=30, face="bold.italic", colour="black"),
    legend.position = c(.10,.85))

  g1 <-
    ggplot(s1, aes(x=s1$Group.1, y=log(s1$x))) +
    geom_line(aes(x=s1$Group.1, y=log(s1$x), colour=log(s1$x)), size=.9) +
    scale_colour_gradient2(low=col1, mid="grey70", high=col2, midpoint = mean
(log(s1$x)), guide=FALSE) +
    xlab("Monthly Frequency") +
    ylab("LOG of Raw Count") +
    ggtitle(ttl) +
    geom_smooth(colour="darkcyan", fill="dimgrey", alpha = .2, size=.8)+
    gtheme

  if(SMOOTH==FALSE){
    g1 <- ggplot(s1, aes(x=s1$Group.1, y=s1$x)) + geom_line(aes(x=s1$Group.1,
y=s1$x), colour=col1, size=.9) +
      xlab("Monthly Frequency") +
      ylab("Raw Count") +
      ggtitle(ttl) +
      gtheme
  }

  g1

}
#]----->
ez.splot(seattle_all, 2)

## `geom_smooth()` using method = 'loess'
```

```
ez.splot_log(seattle_all, 2)

## `geom_smooth()` using method = 'loess'
```

```
#]-----[ CONVERT DATA FRAME LIST TO TIME FRAME LIST ]-----
ez.convert.list <- function(list_tseries_df){
```

```
  name_list <- names(list_tseries_df)
  new <- list()

  for(i in 1:length(list_tseries_df))

    new[[i]] <-
    ts(list_tseries_df[[i]][[2]],
       start=c(as.numeric(format(list_tseries_df[[i]][[1]][1],"%Y")),
               as.numeric(format(list_tseries_df[[i]][[1]][1],"%m"))),
       frequency = 12)

    names(new) <- name_list

    new

  }
#]----->
ts.seattle_all <- ez.convert.list(seattle_all)
#]-----[ REMOVE EXTRA VARIABLE FOR MEMORY ALLOCATION ]-----
rm(
  ez.allday_01,
  ez.convert.list,ez.df_list,
  ez.splot,
  Seattle_Crime_2008_2014
)
#]-----[ SEASONALY ADJUST ORIGINAL TS. ]-----
ez.stl_09 <- function(ts){
  seasadj(stl(ts, 9))
}
#]----->
sa.seattle_all <- lapply(ts.seattle_all, ez.stl_09)
#]-----[ IMPORT ADJUSTED ECON DATA ]-----
load("E:/R_DIRECT_01/R-DIRECT_01/Model_FIN.RData")
#]-----[ REGRESSION MODELING AND RESULTS ]-----
ez.mod_00 <-
  function(n, sa.seattle_all, dfrm_independent){

    name_list <- names(sa.seattle_all)

    mod <-
      lm(as.numeric(log(sa.seattle_all[[n]])) ~
           log(dfrm_independent$cpi) +
           dfrm_independent$unemployment_lag +
           dfrm_independent$econ_growth_index +
           log(dfrm_independent$pop) +
           dfrm_independent$count)

    stargazer(mod, type = "text", title = name_list[n])
```

```
    return(mod)

  }
#]-----------------------------------------< LIST# & NAME>
# 1                Assault
# 2               Burglary
# 7                Robbery
#]-----------------------------------------<>
ez.mod_00(1,sa.seattle_all, dfrm_independent)

##
## Assault
## ========================================================
##                             Dependent variable:
##                      ------------------------------------
##                      as.numeric(log(sa.seattle_all[[n]]))
## --------------------------------------------------------
## cpi)                                2.924
##                                    (2.230)
##
## unemployment_lag                   0.041***
##                                    (0.014)
##
## econ_growth_index                   0.008
##                                    (0.009)
##
## pop)                                -1.949
##                                    (2.654)
##
## count                               0.001
##                                    (0.004)
##
## Constant                            17.043
##                                    (42.411)
##
## --------------------------------------------------------
## Observations                          76
## R2                                  0.405
## Adjusted R2                         0.362
## Residual Std. Error           0.095 (df = 70)
## F Statistic                9.527*** (df = 5; 70)
## ========================================================
## Note:                        *p<0.1; **p<0.05; ***p<0.01

##
## Call:
## lm(formula = as.numeric(log(sa.seattle_all[[n]])) ~ log(dfrm_independent$c
pi) +
##     dfrm_independent$unemployment_lag + dfrm_independent$econ_growth_index
+
```

```
##     log(dfrm_independent$pop) + dfrm_independent$count)
##
## Coefficients:
##                        (Intercept)           log(dfrm_independent$cpi)
##                          17.043241                            2.924146
##  dfrm_independent$unemployment_lag  dfrm_independent$econ_growth_index
##                           0.040999                            0.007961
##          log(dfrm_independent$pop)              dfrm_independent$count
##                          -1.949118                            0.001402
```

```
ez.mod_00(2,sa.seattle_all, dfrm_independent)
```

```
##
## Burglary
## ========================================================
##                              Dependent variable:
##                       ------------------------------------
##                       as.numeric(log(sa.seattle_all[[n]]))
## --------------------------------------------------------
## cpi)                               3.792**
##                                    (1.823)
##
## unemployment_lag                    0.016
##                                    (0.012)
##
## econ_growth_index                  -0.002
##                                    (0.008)
##
## pop)                               -0.539
##                                    (2.170)
##
## count                              -0.002
##                                    (0.003)
##
## Constant                           -6.531
##                                    (34.677)
##
## --------------------------------------------------------
## Observations                          76
## R2                                  0.238
## Adjusted R2                         0.183
## Residual Std. Error           0.078 (df = 70)
## F Statistic                4.367*** (df = 5; 70)
## ========================================================
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Call:
## lm(formula = as.numeric(log(sa.seattle_all[[n]])) ~ log(dfrm_independent$c
pi) +
```

```
##     dfrm_independent$unemployment_lag + dfrm_independent$econ_growth_index
+
##     log(dfrm_independent$pop) + dfrm_independent$count)
##
## Coefficients:
##                        (Intercept)           log(dfrm_independent$cpi)
##                          -6.530851                            3.791648
##  dfrm_independent$unemployment_lag  dfrm_independent$econ_growth_index
##                           0.015582                           -0.001594
##          log(dfrm_independent$pop)              dfrm_independent$count
##                          -0.539362                           -0.002299
```

```
ez.mod_00(7,sa.seattle_all, dfrm_independent)
```

```
##
## Robbery
## ========================================================
##                             Dependent variable:
##                     ------------------------------------
##                     as.numeric(log(sa.seattle_all[[n]]))
## --------------------------------------------------------
## cpi)                             5.057**
##                                  (2.431)
##
## unemployment_lag                  0.007
##                                  (0.016)
##
## econ_growth_index               -0.022**
##                                  (0.010)
##
## pop)                             6.591**
##                                  (2.894)
##
## count                           -0.008**
##                                  (0.004)
##
## Constant                       -117.942**
##                                 (46.242)
##
## --------------------------------------------------------
## Observations                        76
## R2                                0.483
## Adjusted R2                       0.446
## Residual Std. Error          0.104 (df = 70)
## F Statistic               13.079*** (df = 5; 70)
## ========================================================
## Note:                       *p<0.1; **p<0.05; ***p<0.01
```

```
##
## Call:
```

```
## lm(formula = as.numeric(log(sa.seattle_all[[n]])) ~ log(dfrm_independent$c
pi) +
##     dfrm_independent$unemployment_lag + dfrm_independent$econ_growth_index
+
##     log(dfrm_independent$pop) + dfrm_independent$count)
##
## Coefficients:
##                       (Intercept)           log(dfrm_independent$cpi)
##                        -1.179e+02                           5.057e+00
##  dfrm_independent$unemployment_lag  dfrm_independent$econ_growth_index
##                         6.647e-03                          -2.244e-02
##          log(dfrm_independent$pop)              dfrm_independent$count
##                         6.591e+00                          -8.357e-03

#]----------------------------------------------------------[---]
```

Appendix G: Personal Experience

After returning home from my stint in the military I soon came to the realization that I had no marketable skills aside from what I learned as an infantryman. I ended up working private security on the Yakama Reservation for the Yakama Nation Housing YNHA. For those unfamiliar with the reservation, it is only an hour and half drive from CWU (see figure 1 below). The client for whom I worked, YNHA, felt the need for armed security to routinely (every night, 8pm to 5am) patrol their section-8 neighborhoods in order to keep the peace. This was due to the fact that regular law enforcement, including state patrol, was not allowed on these large sections of property due to tribal restrictions. Essentially the only law enforcement with access was Tribal PD, and they had no more than three officers patrolling the entire reservation on any given night. My job entailed working with Tribal PD and providing assistance and coverage in the greater "problem" areas. The effect of brewing such a crime friendly atmosphere was beyond description. I witnessed surges in criminal activities where drive-by shooting would occur multiple times a night. Drug use and prostitution were beyond anyone's control and I routinely saw children who were neglected and uncared for. Home burglaries were so common place, that if someone left their house for more than a day, you could count on it being ransacked in the owner's absence. These are just a few of the crime and social issues I witnessed in these neighborhoods. Throughout my time working this job I was constantly asking the question of why this was occurring here and why criminal activity appeared to increase and decrease so drastically without any apparent cause. Unfortunately, aside from myself, no one seemed to have an answer or even care. And after resolving to find a safer and more meaningful career, I enrolled in CWU and soon became enamored with finding the answers to my many difficult questions using data analysis.

Figure 1