Econometrics of Crime (Naufal Alavi, Luke Sunderman)

Introduction

Sentencing is an element of the United States Justice system that has long been a point of contention. The length of prison sentences in particular have been claimed to be biased for factors to include race, wealth, and others. Team Pantone 448C selected the “crime1” dataset from the wooldridge datasets to investigate this issue in particular. Through the creation, comparison, and interpretation of multiple econometric models, we will search for evidence of such biases. We will not be conducting causal analysis but hope to identify if there are correlations between various factors and the length of prison sentences.

Theoretical Framework/Article Review

The essay “Crime and Punishment: An Economic Approach” (Becker 1968) is a foundational paper that aims to determine how many resources and how much punishment should be used to enforce various laws and regulations. To do this, Becker creates an economic model to determine the elasticities of the total cost of offenses with respect to the probability of conviction and to the magnitude of punishment for those convicted.

The major finding from the paper is that the elasticities of cost to society with respect to the probability of penalization are greater than that of cost to society with respect to severity of punishment. This also makes logical sense, as the cost of punishment is included in the overall cost to society as is the damages from the crime itself. In the instances of more harmful crime, more costly punishment methods such as imprisonment and parole are necessary and thus more costly than the deterrence gained through the improvement of the likelihood of catching an offender.

These findings are not without their limitations however. It is difficult to empirically support them because of severe limitations in available data on convictions and punishment costs. There is also a limitation on the costs of certain offenses. For instance, one individual may take great offense to prostitution and consider it damaging to society while another feels it should be a legal profession.

This paper is relevant to our project because it provides the foundations for the Friesen article on crime compliance and severity and highlights possible limitations on our dataset.

The article “Certainty of Punishment versus Severity of Punishment: An Experimental Investigation" (Friesen 2012) explored the relative effectiveness of equivalent increases in the probability of Punishment and the severity of punishment if caught. Previous findings using general crime data suggested that probability of punishment was more impactful than the severity of punishment but because of the nature of the datasets, studies comparing severity and likelihood of punishment are controversial due to sample selection and endogeneity of crime rate and enforcement parameters.

To remedy this, Friesen conducted an experiment where participants engaged in a simulation where participants would be compensated based on the decisions they make. Participants would be given a starting quantity of money and a cost of compliance. If the participant chose to comply, they would be compensated with their money minus the cost of compliance for participating. They would also be given the option to not comply, which would come with a disclosed risk of audit and cost of infraction. If they chose this option, they would either walk away with their initial cash value if not audited, or they would be compensated with the initial cash minus the cost of compliance.

The general conclusion from this study is that an increase in severity of punishment is a more effective method of compliance than an equivalent increase in the probability of punishment. It is important to note that this finding is limited by the fact that in the real world, a criminal's perception of probability of enforcement and severity is more relevant than the actual probability and severity. Additionally, the pool of participants from the study is students which likely behave differently than the pool of criminals in the real world.

This article is relevant to our project because our independent variable is sentence length, a measure of severity of punishment.

The article “Crime prediction based on crime types and using spatial and temporal criminal hotspots” (Tahani, Mirza, and Lor 2015) attempted to look at data from Denver, Colorado and Los Angeles, California and determine the spatial and temporal hotspots of crime in these cities. The aim was that with information about when and where crime is most likely to happen, citizens would be aided in making better living choices and the police department would be helped with resource allocation. Finally, the paper provided an analysis study by combining the criminal hotspot findings with demographics information, which is the part that becomes relevant to our project. This study concluded that any relationship between crime hotspots and the race distribution of citizens cannot be established. This finding was interesting because we have some information about the race of the convicted people in our $crime1$ dataset for the project. As a result, we included the variables $black$ and $hispan$ in our model to figure out if the race of the convicted felons impact their sentence length. This then extends our research, as having known that there is no correlation between race and crime hotspots, we can analyze whether black and/or hispanic people are given unfair/disproportionate sentences.

All three of these articles relate to each other as a means of establishing effective response to crimes, either by law enforcement agencies or the public at large. Becker 1968 and Friesen 2012 both look at how certainty and severity of punishment impact crime, while Tahani, Mirza, and Lor 2015 focus on finding crime hotspots for effective resource allocation. The ultimate goal is to reduce crime, either by changing certainty and severity of punishment or changing the distribution of resources deployed. Our project and econometric model has a similar goal. By looking at factors impacting the average sentence length for people found guilty of crimes, particularly $qemp86$ and $inc86$, we can say something about whether attempts to increase employment rates and average income could potentially reduce crime.

Using prior literature as a foundation, we constructed our first econometric model to explore possible predictors of sentence length. The econometric model we started our analysis with is as follows:

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 inc86 + \beta_3 black + \beta_4 hispan + u$$

In this model, the response variable $avgsen$ is the average sentence length in months served by the individuals in our dataset from their prior convictions. The first explanatory variable is $qemp86$, which is the number of quarters a person was employed in the year 1986. The second explanatory variable is $inc86$, which is the legal income of each individual in 1986, measured in 100s of dollars. Finally, the last two explanatory variables are $black$ and $hispan$, which are 1 if the person is black or hispanic and 0 if not. All these variables come from the *crime1* dataset in the *wooldridge* package. There are 2725 observations of 16 total variables in this dataset.

Dataset and Descriptive Statistics

A complete legend of all of our variables are as follows:
avgsen: average sentence length in months.
qemp86: number of quarters employed during 1986.
inc86: legal income of each individual during 1986, measured in hundreds of dollars.
black: indicator variable where 1 indicates the person is black and 0 indicates the person is not black.
hispan: indicator variable where 1 indicates the person is hispanic and 0 indicates the person is not hispanic.
pcnv: proportion of prior convictions.

* All variables come from the Wooldridge “Crime1” dataset. Link to the dataset can be found in the citations.

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 inc86 + \beta_3 black + \beta_4 hispan + u$$

Upon running initial regressions, we found that *qemp86* and *inc86* were very closely correlated which could possibly make our model inefficient. To check this, we conducted joint significance F-tests and found *qemp86* and *inc86* to be jointly significant. More details pertaining to hypothesis testing can be found in the final estimation results section. Additional visualizations of our variables are available in Appendix A.

Summary and Descriptive Statistics Tables

Summary Statistics:

Measure	avgsen	qemp86	inc86	pcnv
Mean	0.632	2.309	54.967	0.358
Standard deviation	3.508	1.610	66.627	0.395
Minimum	0	0	0	0
Maximum	59.2	4	541	1

Correlations:

Correlation	avgsen	qemp86	inc86	black	hispan	pcnv
avgsen		-0.107	-0.096	0.119	0.013	0.026
qemp86	-0.107		0.712	-0.159	0.033	-0.004
inc86	-0.096	0.712		-0.147	0.001	-0.009
black	0.119	-0.159	-0.147		-0.231	-0.066
hispan	0.013	0.033	0.001	-0.231		0.010
pcnv	0.026	-0.004	-0.009	-0.066	0.010	

Demographic Totals:

Race	Total
Black	439
Not Black	2286
Hispanic	593
Not Hispanic	2132
Black & Hispanic	0
Not Black & Not Hispanic	1693

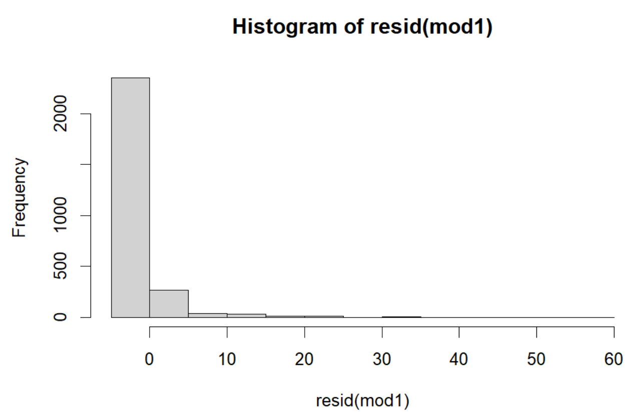
Final Estimation Results

Our base model, called *mod1,* is the same that we reported in the dataset and descriptive statistics section. The fitted values and standard deviations of *mod1* are as follows:

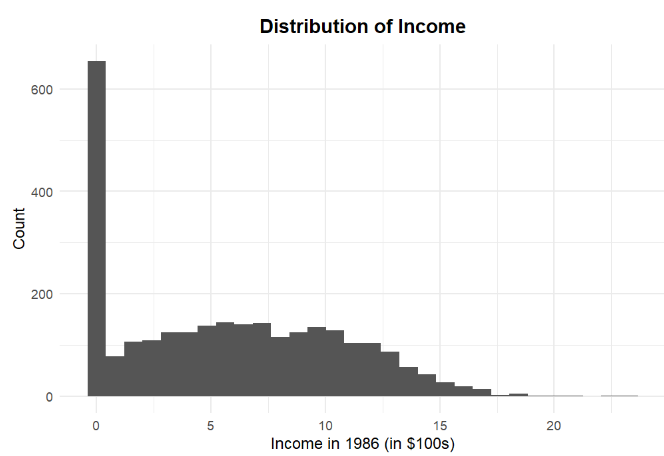
$$\widehat{avgsen} = \underset{(0.1326)}{0.8128} - \underset{(0.0014)}{0.0016}inc86 - \underset{(0.0589)}{0.1485}qemp86 + \underset{(0.1883)}{1.0837}black + \underset{(0.1656)}{0.3511}hispan$$

$$n = 2725,\ R^2 = 0.0243,\ \bar{R}^2 = 0.02286$$

Before we could look at p-values for hypothesis testing, we needed to test our model for homoscedasticity. Upon running the Breusch-Pagan test, we found a BP value of 34.81 with an extremely low p-value of $5.082e^{-7}$. This allows us to conclude that *mod1* is very heteroscedastic and as such we cannot conduct t or F tests using normal standard errors. A histogram of the residuals obtained from *mod1* is shown below.



In an attempt to fix the heteroscedasticity in our model, we decided to use the square root of *inc86,* which was already existent in our original dataset. The figure below shows the distribution of income using the square rooted variable.



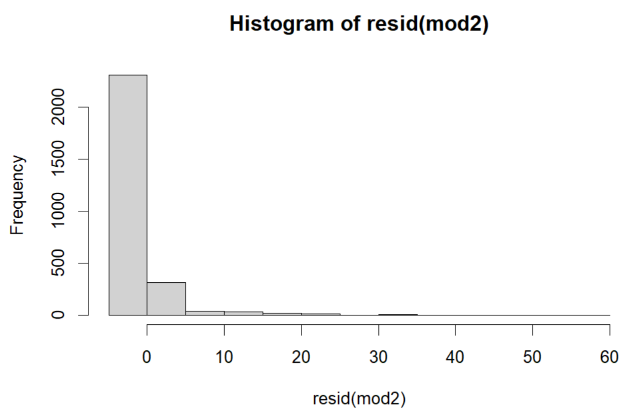
As evident from the figure above, the distribution of income becomes fairly even using the square rooted variable. However, it is heavily skewed to the right because of a large number of instances where income is zero.

Using *sqrt(inc86),* we obtained *mod2* which is as follows:

$$\hat{avgsen} = \underset{(0.1318)}{0.8261} - \underset{(0.029)}{0.0498}\sqrt{inc86} - \underset{(0.0851)}{0.068}qemp86 + \underset{(0.1886)}{1.0691}black + \underset{(0.1657)}{0.3429}hispan$$

$$n = 2725,\ R^2 = 0.0249,\ \bar{R}^2 = 0.02346$$

While the adjusted R^2 value shows a little bit of improvement, upon running the Breusch-Pagan test we found that this model was also heteroscedastic, with a BP value of 35.592 and a p-value of $3.51e^{-7}$. The residuals of *mod2* also weren't too different from those of *mod1.*



In other attempts, we developed two more models, *mod3* and *mod4*, using the same predictors as *mod1* and *mod2* except we used a subset of our dataset with all observations with *inc86* = 0 were omitted. Even with serious data modification, neither of the models were homoscedastic. Additionally we decided it would not be reasonable to restrict our dataset to non-zero values of *inc86* since it would omit roughly half of the observations.

At this stage, we decided to use build models and conduct hypothesis testing using Heteroskedastic Robust standard errors. *Mod1* with robust standard errors follows:

$$\widehat{avgsen} = \underset{\substack{(0.1326)\\ [0.1377]}}{0.8128} - \underset{\substack{(0.0014)\\ [0.0011]}}{0.0016}inc86 - \underset{\substack{(0.0589)\\ [0.0611]}}{0.1485}qemp86 + \underset{\substack{(0.1883)\\ [0.2779]}}{1.0837}black + \underset{\substack{(0.1656)\\ [0.1582]}}{0.3511}hispan$$

$$n = 2725,\ R^2 = 0.0243,\ \bar{R}^2 = 0.02286$$

We then created five more alternative models, called *mod5* through *mod9* respectively.

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 inc86 + \beta_3 black + \beta_4 hispan + \beta_5 pcnv + v$$

$$\widehat{avgsen} = \beta_0 + \beta_1 inc86 + \beta_2 black + \beta_3 hispan + \beta_4 pcnv + w$$

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 black + \beta_3 hispan + \beta_4 pcnv + q$$

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 black + \beta_3 hispan + \beta_4 pcnv + \beta_5 pcnv^2 + e$$

$$\widehat{avgsen} = \beta_0 + \beta_1 qemp86 + \beta_2 black + \beta_3 pcnv + \beta_4 pcnv^2 + r$$

Joint Significance F-Tests Using Heteroskedastic Robust Standard Errors

Since *mod1* and *mod5* (unrestricted) are nested models, we conducted a joint significance F-test to determine which model to keep. The p-value for this test was 0.01287, which suggested that *pcnv* is jointly significant at a 5% significance level, and we must keep it in the regression. Therefore, we picked *mod5* over *mod1*.

Next, we tested *qemp86* for joint significance using models 5 and 6. The p-value for this test was 0.01515, which rejected the null hypothesis at a 5% significance level, prompting us to pick *mod5* once again. However, when we tested *inc86* for joint significance using *mod5* and *mod7,* we derived a p-value of 0.1393. The variable *inc86* was not jointly significant at a 10% significance level, so we ultimately picked *mod7* over *mod5.*

Testing $pcnv^2$ for joint significance, it had an extremely low p-value and was jointly significant. Thus, we picked *mod8* over *mod7.* Finally, we tested *hispan* for joint significance, the result of which was a very high p-value of 0.2531, suggesting that *hispan* was not jointly significant. This conclusion led us to remove *hispan* from the model. As such, our best model turned out to be *mod9.* Shown below is the comparison of all six previously discussed models.

Various Fitted Candidate Models (with Robust Standard Errors)

	Dependent variable:					
	avgsen					
	mod1	mod5	mod6	mod7	mod8	mod9
	(1)	(2)	(3)	(4)	(5)	(6)
inc86	-0.002	-0.002	-0.004***			
	(0.001)	(0.001)	(0.001)			
qemp86	-0.148**	-0.148**		-0.195***	-0.137***	-0.135***
	(0.061)	(0.061)		(0.042)	(0.040)	(0.040)
black	1.084***	1.106***	1.139***	1.118***	1.006***	0.957***
	(0.278)	(0.279)	(0.281)	(0.279)	(0.274)	(0.276)
hispan	0.351**	0.353**	0.341**	0.361**	0.180	
	(0.158)	(0.158)	(0.159)	(0.160)	(0.158)	
pcnv		0.289**	0.290**	0.291**	4.680***	4.781***
		(0.116)	(0.116)	(0.116)	(0.702)	(0.703)
pcnvsq					-4.614***	-4.721***
					(0.715)	(0.716)
Constant	0.813***	0.703***	0.497***	0.719***	0.383***	0.422***
	(0.138)	(0.147)	(0.098)	(0.144)	(0.133)	(0.131)
Observations	2,725	2,725	2,725	2,725	2,725	2,725
R2	0.024	0.025	0.023	0.025	0.043	0.043
Adjusted R2	0.023	0.024	0.022	0.023	0.041	0.041
Note:				*p<0.1; **p<0.05; ***p<0.01		

Final Model Results

Our final model, *mod9,* is as follows:

$$\widehat{avgsen} = 0.422 - 0.135qemp86 + 0.957black + 4.781pcnv - 4.721pcnv^2$$
$$\begin{array}{ccccc} (0.148) & (0.042) & (0.182) & (0.629) & (0.637) \\ [0.131] & [0.040] & [0.276] & [0.703] & [0.716] \end{array}$$

$$n = 2725,\ R^2 = 0.04251,\ \bar{R}^2 = 0.0411$$

Statistical Significance of Estimated Coefficients and Overall Significance

As seen from the Stargazer table above, all of the predictors in *mod9* are individually statistically significant at a 1% significance level. This means that all of our predictor variables have a statistically significant effect on the average length of sentences served. We have an R^2 value of 0.04251, which suggests that all of our predictor variables together explain 4.251% of the variation in *avgsen.* With an adjusted R^2 value of 0.04251, *mod9* is the best possible model we could build with our dataset. Testing for overall significance, we get an F-value of 15.172 and a low p-value of $2.842e^{-12}$, which suggests that all of our predictors together are statistically significant. Interpretations of the individual predictors in this model are detailed in the following sections.

Quarters Employed in 1986

Holding all other factors constant, we expect an increase in the number of quarters employed in 1986 by one to decrease the average sentence served by a person by 0.135 months. This makes sense as the more time a person spends employed, the more money they might be making which could discourage them to engage in criminal activities, and serve shorter sentences on average. More importantly, employment is time consuming, leaving less time to engage in such activities regardless of the amount of money they earn. Depending on the wages, some individuals may have earned less money than others despite being employed for longer periods. They will still, however, have less time to spare leading to shorter sentences on average.

Black

Holding all other factors constant, we expect a black individual to serve an average sentence of 0.957 months more than a non-black individual. This is a very interesting finding with serious implications. As mentioned in our literature review section, any relationship between crime hotspots and the race distribution of citizens cannot be established (Tahani, Mirza, and Lor 2015). Despite this, there is statistically significant evidence that being black increases the length of the average sentence served. This might imply the existence of disproportionate sentencing of black individuals. However, we need to consider the fact that the data for both studies was collected more than twenty years apart. It is quite possible for either of the conclusions to be different depending on when and where the observations were made. Moreover, the data in the aforementioned paper was collected only from two cities, Denver and Los Angeles. The paper's finding that there is no connection between crime hotspots and race is in line with our conclusion that being hispanic does not significantly impact average sentence lengths as shown by the tests conducted in the previous section.

Diminishing Marginal Effect of Proportion of Prior Convictions on Average Sentence

$$\frac{\partial avgsen}{\partial pcnv} = 4.781 - 9.442pcnv$$
$$\frac{\partial avgsen}{\partial pcnv} = 0$$
$$4.781 - 9.442pcnv = 0$$
$$4.781 = 9.442pcnv$$
$$pcnv = 0.506$$

The turning point is 0.506. This implies that after a *pcnv* value of 0.506 (all else held constant), the marginal effect of proportion of prior convictions on average sentence becomes negative and *pcnv* starts to decrease *avgsen.*

Conclusion

To conclude, we have evidence that the number of quarters employed plays a prominent role in reducing the average sentence length served by an individual. Moreover, a black individual will on average serve a sentence that is longer by almost an entire month than a non-black individual. However, whether or not there is a disproportionate incarceration of black individuals with respect to the amount of crime they commit requires further research. Proportion of prior convictions is a factor in increasing average sentence lengths, but it has a diminishing marginal effect after a little over 50%. It is, however, difficult to empirically support these findings because of the limited available data on convictions, which was an issue that was forecasted for us in “Crime and Punishment: An Economic Approach” (Becker 1968). More data needs to be studied to provide any conclusive evidence to our findings. Moreover, even though our estimations are statistically significant, whether or not they are economically significant is arguable, since both *qemp86* and *black* impact *avgsen* by only 4 and 29 days respectively.

Citations and Weblinks

Almanie Tahani, Rsha Mirza, and Elizabeth Lor. "Crime prediction based on crime types and using spatial and temporal criminal hotspots." *International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.5,* No.4 (2015): 1-20.

Link: https://arxiv.org/ftp/arxiv/papers/1508/1508.02050.pdf

Friesen, Lana. “Certainty of Punishment versus Severity of Punishment: An Experimental Investigation.” *Southern Economic Journal* 79, no. 2 (2012): 399–421.

Link: http://www.jstor.org/stable/41638882

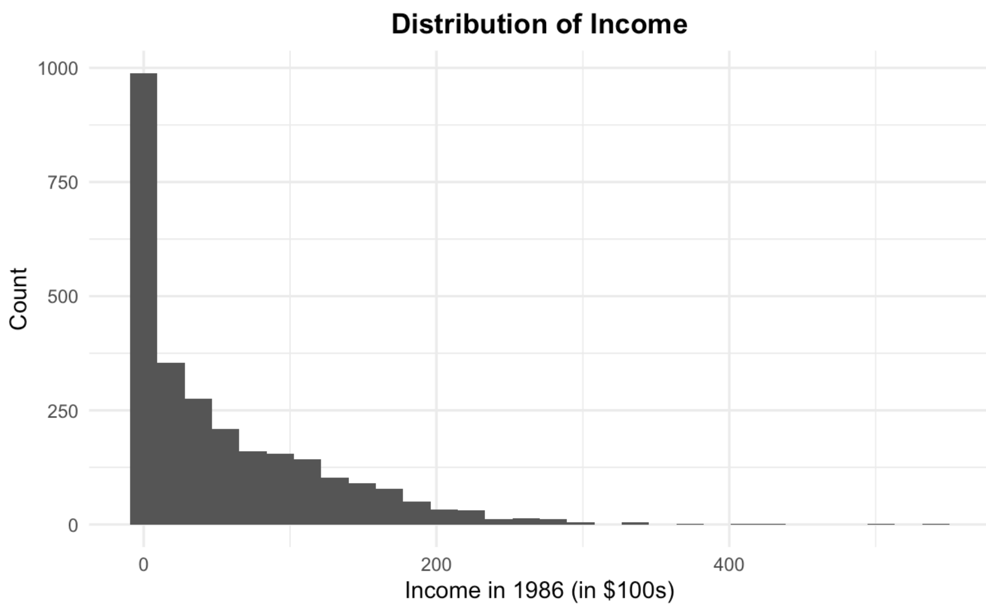
Becker, Gary S. “Crime and Punishment: An Economic Approach.” *Journal of Political Economy* 76, no. 2 (1968): 169–217.

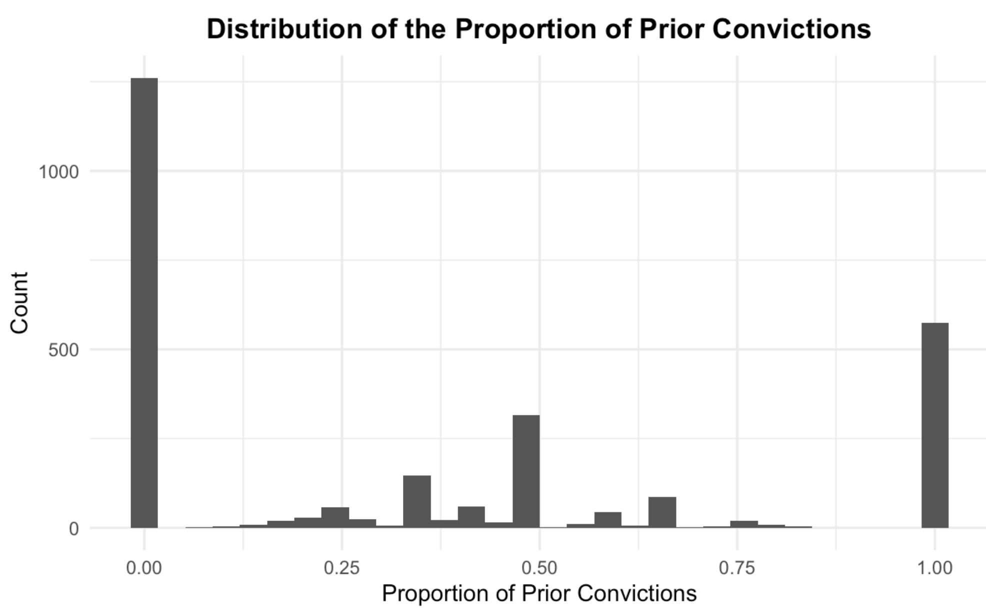
Link: http://www.jstor.org/stable/1830482

Dataset source:
https://www.cengage.com/cgi-wadsworth/course_products_wp.pl?fid=M20b&product_isbn_issn=9781111531041

Appendix A

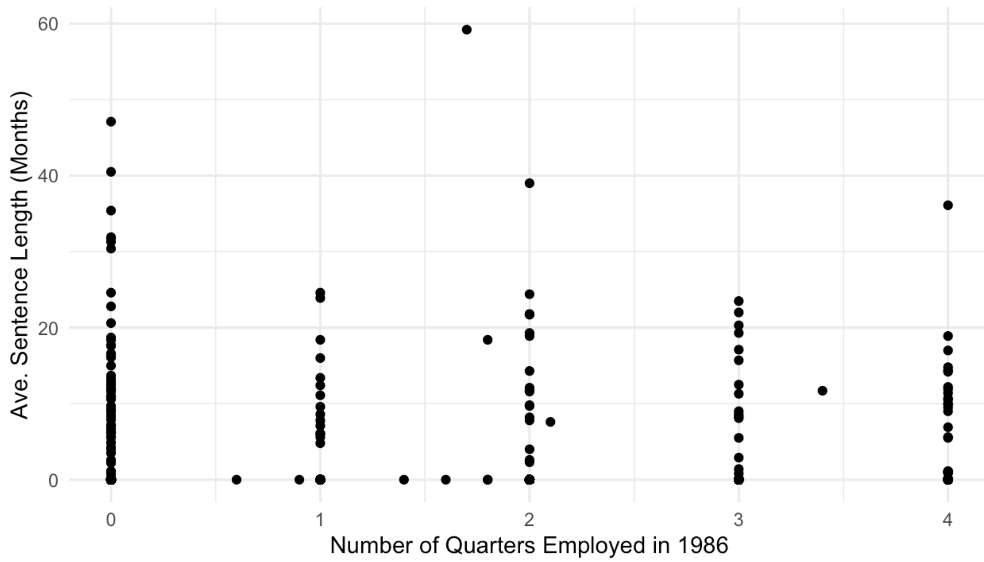

Distribution of Employment
Count
1000
750
500
250
0
0
1
2
3
4
Number of Quarters Employed in 1986


Distribution of Income
Count
1000
750
500
250
0
0
200
400
Income in 1986 (in $100s)

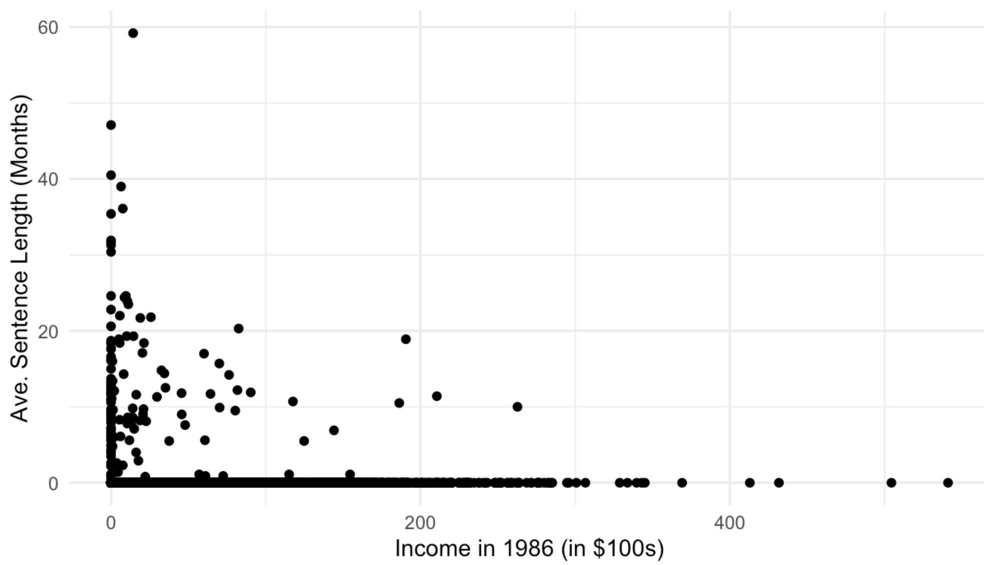

Distribution of the Proportion of Prior Convictions
Count
1000
500
0
0.00
0.25
0.50
0.75
1.00
Proportion of Prior Convictions

Appendix A (Continued)

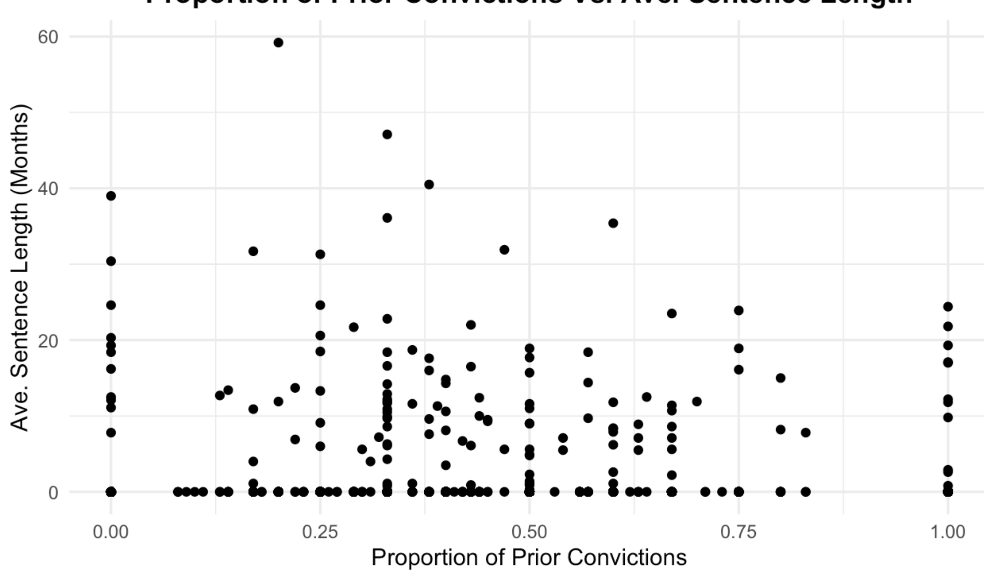
Employment Vs. Ave. Sentence Length



Income Vs. Ave. Sentence Length



Proportion of Prior Convictions Vs. Ave. Sentence Length



Appendix A (Continued)