Prediction of Property Crimes in Georgia

Submitted by
Luiz Gustavo Fagundes Malpele
Department of Data Science
Florida Polytechnic University

December 6th, 2019

Abstract: Georgia's state government sought to obtain predictions of property crime rates by county for the purpose of state level law enforcement budgeting and planning. In order to accomplish this, the property crime rate was predicted by utilizing past crime rates, growth rates, and population-related factors. Then, different models were generated utilizing socioeconomic factors, and the best one was selected. Finally, a combination of historical crime rates and the best socioeconomic factors were utilized to see if the models produced had better predictive power. The results were that property crimes in the state of Georgia were better predicted, utilizing both historical factors related to property crimes and socioeconomic factors related to the population's poverty level.

1. Introduction

The Georgia state government desires predictions of crime levels by county for state law enforcement budge and planning. By knowing the most impacting factors that increase the property crime rates, it is possible to allocate resources to solve the problem at its source. This analysis also provides data regarding the expected amount of property crimes to the year 2019 and based on this information. Each county can adequately prepare for the upcoming year.

For such prediction to be made, the best models were developed using the historical data, which included data of past crime rates and population-related factor and socioeconomic-related factors, such as poverty, education, unemployment, and age. Finally, the best models for each of the historical and socioeconomic factors were combined with the objective of increasing the predictive power, and a prediction will be made based off on the best overall models from the analysis.

2. Data

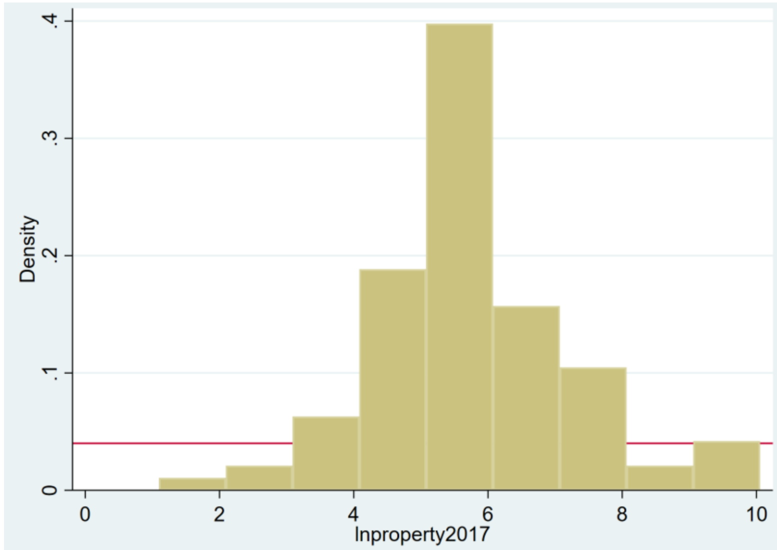
The data utilized in this project came from the Federal Bureau of Investigation's Uniform Crime Reporting Program, which provided data related to the number of property and violent crimes grouped by county for the years of 2012, 2017, and 2018. The second utilized data source was the Census Bureau's American Fact Finder, which provides access to data about United States, Puerto Rico, and the Island Areas acquired from several censuses and surveys. Furthermore, it was utilized from this second data source specific social and economic factors such as age breakdown, education, unemployment, and poverty for each county within Georgia.

It is important to note that the data gathered from these two sources was wrangled and cleaned to be able to properly format it and investigate the best regression models that would best predict the property crimes in county within Georgia. A summary and description of all the variables can be found in Table 2.

Lastly, the top 4 counties inflated the RMSE values of the prediction model as the number of property crimes in the year of 2017 on Clayton, Cobb, DeKalb, and Gwinnett were 8711, 11899, 23246, and 15079 respectively. This fact caused the top 4 percentile to increase the variance of property2017 variable to 2877.916, which is a high value that inflated the RMSE values of the models.

The distribution of the lnproperty2017 is displayed below (Table 1) and it can be observed that this variable is normally distributed since the curve is bell-shaped, but there is substantial an amount of counties on the last bin, these represent 0.04 on the density axis, which are the 4 previously mentioned counties (Clayton, Cobb, DeKalb, and Gwinnett).

Tables 1: Frequency distribution of the lnproperty2017 variable



Tables 2: Summary of the variables (Mean, Standard deviation, Minimum, and Maximum)

Variable	Mean	Std. Dev.	Minimum	Maximum
violent2012	119.8021	402.174	0	3420
violent2017	122.5208	388.4895	0	3184
violent2018	128.8438	389.6356	0	3085
violentgrowth	0.0291374	0.1216718	-0.3531568	0.2834132
lnviolent2012	3.489516	1.394921	0	8.137396
lnviolent2017	3.626937	1.279686	0.6931472	8.065893
property2012	1271.885	3672.148	2	30285
property2017	1109.042	3161.136	3	23246
property2018	978	2877.916	1	21426
popgrowth	0.0033757	0.0107121	-0.013552	0.0697667
lnpop2012	10.28049	1.192467	7.925158	13.78838
lnpop2017	10.29737	1.221712	7.861727	13.85366
pop2012	74038.13	163019.8	2766	973236
pop2017	78400.6	175426.6	2596	1038884
propertygrowth	0.0349247	0.0944389	0.5278115	0.5051457
lnproperty2012	5.888458	1.467278	0.6931472	10.31841
lnproperty2017	5.713834	1.444298	1.098612	10.05389
poppov	74895.93	11167.23	187	77387
poor	6569.26	13266.33	322	79132

verypoor	5485.75	11167.23	187	77387
unemployment	8.20625	3.004481	4	20.3
eduless9th	5.896875	2.476859	1.5	14
edu9to12th	11.85208	3.774289	3.2	22.8
highschool	35.37396	6.50032	18.3	52.1
associate	7.396875	1.521388	3.4	12.2
somecollege	20.58437	3.261234	2.1	25.2
bachelor	11.43646	5.627566	3	29.9
graduatedegree	7.464583	3.789653	2.1	25.2
pop14to172012	4176.781	9046.742	119	54311
pop14to172017	4445.188	9903.134	149	60463
pop18to242012	7235.083	16102.45	216	105314
pop18to242017	7289.813	16524.48	190	105701
pop65up2012	8192.01	14809.5	412	93402
pop65up2017	10273.48	19364.88	493	118228
pop85up2012	881.0833	1755.629	33	12704
pop65to842017	9245.125	17337.7	447	104127
share14to17	.0540882	.0076406	.0325851	.07153
share18to24	.0883316	.0250651	.0592906	.2374612
Share65to84	.1551222	.0426909	.0784637	.306959
share85up	.0177731	.0063023	.0063191	.0387641
lessthanhs	17.74375	5.759857	5.000001	31.4
shareold	.1728953	.04791	.0880881	.3445706
sharepoor	.1133421	.0386926	.037361	.2119249
sharepoverty	.2069274	.0694555	.0630627	.4144205
shareverypoor	.0935853	.0401298	.0224568	.2611427
shareyoung	.1424199	.0248166	.100813	.2803217
lessthanhs	17.74375	5.759857	5.000001	31.4
bachelorup	18.90104	9.120185	6.9	50.2

The number of observations for each variable is equal to 96, the variables summary shows a trend of population growth and a decrease on property crime occurrences with a large standard deviation value for variables that are not log-transformed or do not refer to a percentage value, there is a significant number of highly-populated counties and this fact makes the mean and standard deviation value significantly increase. The data will be further analyzed using the Poisson regression model as it is the best one to work with count data, and property crime rates are a good example of it, since it is not an everyday event, and the values were captured in a yearly-based period.

Table 3: Description of all variable (Storage type, Display format, Variable label)

Variable	Storage type	Display format	Variable label
violent2012	int	%8.0g	Violent2012
violent2017	int	%8.0g	Violent2017
violent2018	int	%8.0g	Violent2018
violentgrowth	float	%9.0g	
lnviolent2012	float	%9.0g	
lnviolent2017	float	%9.0g	
property2012	int	%8.0g	Property2012
property2017	int	%8.0g	Property2017
property2018	int	%8.0g	Property2018
popgrowth	float	%9.0g	
lnpop2012	float	%9.0g	
lnpop2017	float	%12.0g	
pop2012	long	%12.0g	Pop2012
pop2017	long	%12.0g	Pop2017
propertygrowth	float	%9.0g	
lnproperty2012	float	%9.0g	
lnproperty2017	float	%9.0g	
poppov	long	%12.0g	PopPov
poor	long	%12.0g	Poor
verypoor	long	%12.0g	VeryPoor
unemployment	float	%9.0g	Unemployment
eduless9th	float	%9.0g	Eduless9th
edu9to12th	float	%9.0g	Edu9to12th
highschool	float	%9.0g	Highschool
associate	float	%9.0g	Associate
somecollege	float	%9.0g	SomeCollege
bachelor	float	%9.0g	Bachelor
graduatedegree	float	%9.0g	Graduate degree
pop14to172012	long	%12.0g	Pop14to172012
pop14to172017	long	%12.0g	Pop14to172017
pop18to242012	long	%12.0g	Pop18to242012
pop18to242017	long	%12.0g	Pop18to242017
pop65up2012	long	%12.0g	Pop65up2012
pop65up2017	long	%12.0g	Pop65up2017
pop85up2012	int	%8.0g	Pop85up2012

pop85up2017	int	%8.0g	Pop85up2017
pop65to842017	float	%9.0g	
share14to17	float	%9.0g	
share18to24	float	%9.0g	
share65to84	float	%9.0g	
share85up	float	%9.0g	
shareold	float	%9.0g	
sharepoor	float	%9.0g	
sharepoverty	float	%9.0g	
shareverypoor	float	%9.0g	
shareyoung	float	%9.0g	
bachelorup	float	%9.0g	
lessthanhs	float	%9.0g	

This table includes information regarding the storage type of each variable (int, float, or long), the display format for each of the variables, and the variable label, which is the variable name on the Excel file before importing it to Stata.

3.1 Analysis – Prediction Based on Past Crimes and Population Factors

Table 4: Correlation matrix for property, violence, and population data.

	pro~2018	lnvio~2017	viole~h	lnpr~2017	pr~growth	lnpop2017	popgrowth
property2018	1.0000						
lnviole~2017	0.5904	1.0000					
violentgro~h	-0.0877	0.0578	1.0000				
lnprope~2017	0.6762	0.7736	-0.2816	1.0000			
propertygr~h	0.0271	-0.0321	0.3266	-0.0654	1.0000		
lnpop2017	0.6171	0.7295	-0.1841	0.8802	-0.1026	1.0000	
popgrowth	0.2623	0.3206	-0.1024	0.4586	-0.0202	0.5404	1.0000

The correlation matrix indicates the degrees of collinearity of the variables; when the absolute value of the Pearson's r is above 0.5, it means there is a high a degree of collinearity between the variables and that one of those two variables are more likely to be statistically significant for the analysis.

The relations of *lnviolent2017* and *property2018* have r value equals to 0.5904, which means that these two variables are highly correlated; the same case occurs for *lnproperty2017* and *property2018* (r = 0.6762). These two cases refer to crime data, and the reason for these occurrences can be described by the use of similar data, as there was not much difference in crime rates between consecutive years, and there might not be much difference between property and violent crime rates.

There is strong evidence that the *lnpop2017* variable should not be used on the analysis since there are high degree of collinearity of this variable with 3 other variables: *lnpropertygrowth* and *lnviolence2017* (r = 0.7736), *lnpop2017* and *property2018* (r = 0.6171), *lnpop2017* and *lnviolent2017* (r = 0.7295), *lnpop2017* and *lnproperty2017* (r = 0.8802).

The high collinearity case of *popgrowth* and *lnpop2017* (r = 0.5404) can be explained by the fact that both are population variables and *populationgrowt* involves the *lnpop2017* to be calculated.

Table 5: Robust-Poisson regression of lnviolent2017, violentgrowth, lnproperty2017, propertygrowth, lnpop2017, popgrowth on property2018 (6 predictor)

Poisson regression		Number of obs =	92
		Wald chi2(6) =	15617.55
		Prob > chi2 =	0.0000
Log pseudolikelihood =	-2111.8536	Pseudo R2 =	0.9841

property2018	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnviolent2017	-.0059011	.0109741	-0.54	0.591	-.0274099	.0156077
violentgrowth	.164006	.2144143	0.76	0.444	-.2562382	.5842503
lnproperty2017	1.106563	.0980341	11.29	0.000	.9144202	1.298707
propertygrowth	-.3762712	.6124239	-0.61	0.539	-1.5766	.8240576
lnpop2017	-.1453132	.1426851	-1.02	0.308	-.4249709	.1343446
popgrowth	7.329432	5.282404	1.39	0.165	-3.023889	17.68275
_cons	.7267536	.8546479	0.85	0.395	-.9483255	2.401833

Poisson regression was used to analyze the effects of 6 different predictor predictors on the response variable *property2018*. It is noticeable that five different variables, which are *lnviolent2017*, *violentgrowth*, *propertygrowth*, *popgrowth*, and *lnpop2017*, have a p-value above 0.05, which indicates that it is necessary to check through joint tests for the statistical significance of these variables on *property2018*. The next step is to compare RMSE values for different models and the joint hypotheses tests, then, it will be possible to choose the one that accounts for the best predictors for the analysis of property crimes in the year of 2018.

Table 6: Chi-Square test for the effect of lnpop2017 and popgrowth on property2018

```
( 1)  [property2018]lnpop2017 = 0
( 2)  [property2018]popgrowth = 0

           chi2(  2) =    1.97
         Prob > chi2 =    0.3728
```

Given that the p-value equals 0.3728 for the joint effect of lnpop2017 and popgrowth and α = 0.05, there is not enough strong evidence to state that at least one of these variables is equal to zero. This fact indicates that these two population variables may not be statistically significant to the analysis of property crimes in the state of Georgia.

Table 7: Chi-Square test for the effect of lnpop2017 and popgrowth on property2018

```
( 1)  [property2018]lnviolent2017 = 0
( 2)  [property2018]violentgrowth = 0

           chi2(  2) =    0.59
         Prob > chi2 =    0.7455
```

Given that the p-value equals to 0.7455 for the joint effect of *lnviolent2017* and *violentgrowth* and α = 0.05, so there is not enough strong evidence to state that at least one of these variables is equal to zero and strong evidence that at least one the β coefficients is equal 0. This fact indicates that these two population variables may not be statistically significant to the analysis of property crimes in the state of Georgia.

Table 8: Robust-Poisson regression of lnviolent2017, violentgrowth, lnproperty2017, propertygrowth, lnpop2017, popgrowth on property2018 (4 predictor)

```
Poisson regression                              Number of obs     =         92
                                                Wald chi2(4)      =   11066.89
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -2332.6422               Pseudo R2         =     0.9824
```

property2018	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnviolent2017	-.022071	.0138086	-1.60	0.110	-.0491354	.0049933
violentgrowth	.0614741	.3090819	0.20	0.842	-.5443154	.6672635
lnproperty2017	1.019412	.0116899	87.20	0.000	.9964999	1.042324
propertygrowth	-.5839831	.5853319	-1.00	0.318	-1.731213	.5632464
_cons	-.1763327	.0845377	-2.09	0.037	-.3420235	-.0106419

Since the p-values for lnpop2017 accounted for a high degree of collinearity with three other variables and the joint test of *lnpop2017* and *popgrowth* indicated that these two variables were not statistically significant for the analysis, they were removed from the Poisson model and it was performed with four predictors: *lnviolent2017*, *violentgrowth*, *lnproperty2017*, *propertygrowth*.

In the case of this new Poisson model (Table 8), there are still inconsistencies since the p-value for *violentgrowth* is equal 0.842, so there is strong evidence that this variable is equal to 0 and does not affect the prediction of property crimes — the next step to check this fact by performing joint tests.

Table 9: Chi-Square test for the effect of lnviolent2017 and violentgrowth on property2018

```
 ( 1)  [property2018]lnviolent2017 = 0
 ( 2)  [property2018]violentgrowth = 0

           chi2(  2) =    4.44
         Prob > chi2 =    0.1089
```

Given that the p-value equals to 0.1089 for the joint effect of *lnviolent2017* and *violentgrowth* and $\alpha = 0.05$, so there is not enough strong evidence to state that at least one of these variables is equal to zero and strong evidence that at least one the β coefficients equals 0. This fact indicates that these two population variables may not be statistically significant to the analysis of property crimes in the state of Georgia and that a new model with different variables is necessary.

Table 10: Robust-Poisson regression of lnproperty2017, propertygrowth on property2018 (2 predictor)

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(2)      =   11022.00
                                                Prob > chi2       =     0.0000
Log pseudolikelihood =  -2398.292               Pseudo R2         =     0.9823

------------------------------------------------------------------------------
             |               Robust
property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
propertygrowth |   -.504605   .4927611    -1.02   0.306    -1.470399    .461189
lnproperty2017 |    .997609   .0103021    96.84   0.000     .9774172   1.017801
       _cons |  -.1176997   .0754387    -1.56   0.119    -.2655569    .0301574
------------------------------------------------------------------------------
```

This third model (Table 10) presents a consistent result since the individual coefficients of propertygrowth and lnproperty2017 have a p-value of 0.306 and 0.000 respectively, and their joint test yields a p-value equals to 0.000. Furthermore, the Pearson's r for the correlation of the 3 variables involved the correlation (Table 11) are reasonable, with the exception of the relation of lnproperty2017 and property2018, since it presents a Pearson's r equals to 0.6455, which indicates high degree of collinearity of between the property2018 and lnproperty2017, but this is justified by the fact that these two variables are highly correlated and they refer to the property crime data, one in the year of 2017 and the other 2018. The correlation matrix (Table 11) shows each of the relations:

3.2 Model Selection for Historical Factors

Table 11: Correlation matrix for property, violence, and population data.

	property2018	**propertygrowth**	**lnproperty2017**
property2018	1.0000		
propertygrowth	0.0382	1.0000	
lnproperty2017	0.6455	0.1144	1.0000

Table 12: RMSE comparison of the three models

	(1) property2018	(2) property2018	(3) property2018
property2018			
lnviolent2017	-0.00586	-0.0221	
	(0.011)	(0.014)	
violentgrowth	0.161	0.0614	
lnproperty2017	1.107***	1.019***	0.998***
	(0.098)	(0.012)	(0.010)
propertygrowth	-0.371	-0.584	-0.505
	(0.619)	(0.589)	(0.494)
lnpop2017	-0.146		
	(0.143)		
popgrowth	7.350		
	(5.309)		
_cons	0.728	-0.176*	-0.118
	(0.856)	(0.085)	(0.076)
RMSE	697.879	754.372	468.806
N	91	91	95

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

By taking into consideration the RMSE values, joint tests, and statistically significance of each coefficient, it is reasonable to state that model 3 with lnproperty2017 and *propertygrowth* as the coefficients is the most suitable regression model, the RMSE value is the best and the predictors are significantly more statistically significant while compared to the other 2 tested models.

3.3 Multicollinearity on socioeconomic parameters

The next step for the model selection is to investigate socio-economic variables that are useful for the prediction of property crimes in Georgia. This way, the correlation matrix indicates the degrees of collinearity of the variables; when the absolute value of the Pearson's r is above 0.5, it means there is a high a degree of collinearity between the variables and that one of those two variables are more likely to be statistically significant for the analysis. For the matrix below (Table 13), nine different socio-economic variables were taken into consideration for this analysis which are: *property2017*, *lnpop2017*, *popgrowth*, *shareyoung*, *shareold*, *unemployment*, *sharepoverty*, *lessthanhs*, *bachelorup*.

Table 13: Correlation matrix for 9 different socio-economic variables

	property2017	lnpop2017	popgrowth	shareyoung	shareold	unemployment	sharepoverty	lessthanhs	bachelorup
property2017	1.0000								
lnpop2017	0.6383	1.0000							
popgrowth	0.2753	0.5614	1.0000						
shareyoung	0.0622	0.3219	0.0777	1.0000					
shareold	-0.3345	-0.5563	-0.2526	-0.4963	1.0000				
unemployment	-0.0782	-0.1848	-0.2512	0.0863	-0.0468	1.0000			
sharepoverty	-0.2184	-0.5184	-0.6938	0.0260	0.1813	0.4265	1.0000		
lessthanhs	-0.2814	-0.5495	-0.6068	-0.2463	0.2130	0.0719	0.6386	1.0000	
bachelorup	0.5103	0.7198	0.6480	0.2251	-0.2502	-0.2580	-0.6269	-0.7673	1.000

From the correlation matrix above it possible to drive some insight based on the Person's r for the relation of each of the coefficients. The *bachelorup* variable is highly correlated with five of the eight other analyzed variables which are *property2017* (r = 0.5103), *lnpop2017* (0.7198), and *popgrowth* (r = 0.6480), the reason for these is because in higher populated counties it is more likely to have citizens with a bachelor or superior degree, and it is negatively correlated with *sharepoverty* (r = -0.6269), *lessthanhs* (r = -0.7673) and the assumptions are intuitive since where poverty is high, the county is more likely to have less graduates and the same thing can be said for the *lessthanhs* variable.

The *lessthanhs* variable is highly correlated with 3 other variables which are *sharepoverty* (r = 0.6386), *lnpop2017* (r = -0.6068), *popgrowth* (r = -0.5495), these last two can be explained by the fact that for growing and highly-populated counties, students are more likely to get their high school diploma and the correlation with *sharepoverty* is due to the fact that in poverty is related to lower-levels of education.

Lastly, lnpop2017 presented a high degree of correlation with 5 other variables (*popgrwoth*, *shareold*, *sharepoverty*, *lessthanhs*, and *bachelorup*) and *popgrowth* with 3 other variables (*sharepoverty*, *lessthanhs*, *bachelorup*). Since, the *lnpop* variable will be used on the regression models, *popgrowth* with more tests and evidence could be removed from the models if valid.

Table 14: VIF for socio-economic variables

Variable	VIF	1/VIF
bachelorup	4.07	0.245550
lessthanhs	3.30	0.303298
lnpop2017	3.12	0.320822
sharepoverty	3.02	0.331493
popgrowth	2.34	0.428202
shareold	1.93	0.517072
shareyoung	1.53	0.651540
unemployment	1.43	0.698171
Mean VIF	2.59	

This Variance inflation factor list provides useful insight and reflects the cases of collinearity point out previously. The *bacherlorup* variable has a VIF equal to 4.07 which is an indication that this variable may not be suitable for the analysis as the variance of coefficients estimates is over 4.07 times larger due to multicollinearity, the RMSE values need to be evaluated to drive to a consistent conclusion if this variable is removed or not.

3.4 Parameters Replacement and Hypothesis Testing

After understanding the coefficients that are going to be used in the analysis and check for their correlation, a baseline model for the analysis (Table 14) is created. The Robust-Poisson model was chosen because since the data can be described as count data as the counts have been made within a fixed period, the year 2017, for this case.

A baseline model was created, so it is possible to evaluate how replacing the coefficients change the prediction power while comparing the results to the baseline model. The socioeconomic variable that are going to be used on this first model are: *lnpop2017, popgrowth, shareyoung, unemployment, sharepoverty, lessthanhs,* and *bachelorup.*

Table 15: Robust-Poisson regression baseline model with all variables

Poisson regression	Number of obs = 96
	Wald chi2(8) = 768.49
	Prob > chi2 = 0.0000
Log pseudolikelihood = -15076.717	Pseudo R2 = 0.9011

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.318093	.1512191	8.72	0.000	1.021709	1.614477
popgrowth	3.042099	14.39551	0.21	0.833	-25.17258	31.25678
shareyoung	-21.69349	12.29053	-1.77	0.078	-45.78247	2.395501
shareold	-4.202415	3.007807	-1.40	0.162	-10.09761	1.692777
unemployment	.0147603	.0356234	0.41	0.679	-.0550603	.0845809
sharepoverty	3.259106	3.537914	0.92	0.357	-3.675078	10.19329
lessthanhs	-.0096915	.0197519	-0.49	0.624	-.0484045	.0290215
bachelorup	-.0328026	.0256176	-1.28	0.200	-.0830121	.017407
_cons	-3.954115	2.013414	-1.96	0.050	-7.900335	-.0078954

Taking into consideration that the RMSE for this model is equal to 2456.76, which is a high value when compared to other models, and the presence of some variables with the problem of high degree of collinearity such as *bachelhorup* (VIF = 4.07), there are improvements that can be performed in order to make better predictions of property crimes.

Table 16: Robust-Poisson regression replacing shareyoung with share14to17 and share18to24

Poisson regression	Number of obs = 96
	Wald chi2(9) = 653.97
	Prob > chi2 = 0.0000
Log pseudolikelihood = -14678.947	Pseudo R2 = 0.9037

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.335356	.1607283	8.31	0.000	1.020334	1.650378
popgrowth	5.546575	14.1491	0.39	0.695	-22.18514	33.27829
share14to17	-43.5058	20.71389	-2.10	0.036	-84.10429	-2.907312
share18to24	-18.51355	9.526645	-1.94	0.052	-37.18543	.158328
shareold	-5.65906	3.053151	-1.85	0.064	-11.64313	.3250053
unemployment	.0031606	.0339404	0.09	0.926	-.0633614	.0696827
sharepoverty	1.521039	3.434404	0.44	0.658	-5.210269	8.252347
lessthanhs	-.0097676	.0200316	-0.49	0.626	-.0490287	.0294936
bachelorup	-.0444096	.0269683	-1.65	0.100	-.0972666	.0084473
_cons	-2.326547	2.078942	-1.12	0.263	-6.401197	1.748104

The RMSE value equals 2446.43, which is a lower value than the base model (RMSE = 2456.76), suggesting that sorting the *shareyoung* variable into *share14to17* and *share18to24* is a better approach in order to get the best fit for the Poisson regression line. The next step is to analyze the p-values of replaced coefficients.

Table 17: Chi-Square test for the joint effect of share18to24 and share14to17 on property2017

```
( 1)  [property2017]share18to24 = 0
( 2)  [property2017]share14to17 = 0

           chi2(  2) =    5.88
         Prob > chi2 =    0.0528
```

Given that the p-value of the Chi-Square test for the joint effect of *share65to84* and *share 89up* equals to 0.0528 and that α = 0.05, there is moderate evidence to state that at least one of coefficients of these two variables should not be different from zero. Furthermore, the RMSE value of this model is 10.3 units lower than the baseline model, which is not a statistically significant number, but it supports the argument that replacing these the *shareyoung* variable would be a better approach to the analysis, this variable had a p-value of 0.0780 on the baseline model which is 0.0258 higher than the model with the replaced variables.

Table 18: Chi-Square test comparison of coefficients of variables share18to24 and share14to17

```
( 1)  - [property2017]share14to17 + [property2017]share18to24 = 0

           chi2(  1) =    1.68
         Prob > chi2 =    0.1947
```

The comparison of *share14to17* and *share18to24* does not support the argument that these two variables should not replace *shareyoung* since the p-value of the Chi-Square test is 0.1947, the result of this test is inconclusive since the p-value is above 0.05. Indeed, since the previous analyzed evidence has shown that the replacement variable does not yield not statistically significant changes to the model, the best approach is not replacing the shareyoung variable.

Table 19: Robust-Poisson regression replacing shareold with share65to84 and share85up

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(9)      =    1282.20
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -15075.86               Pseudo R2         =     0.9011
```

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.317202	.1421603	9.27	0.000	1.038573	1.595831
popgrowth	2.834533	17.93098	0.16	0.874	-32.30954	37.9786
shareyoung	-21.68312	12.20048	-1.78	0.076	-45.59563	2.229386
share65to84	-3.912239	8.41269	-0.47	0.642	-20.40081	12.57633
share85up	-6.695603	74.94257	-0.09	0.929	-153.5803	140.1891
unemployment	.0143257	.0314011	0.46	0.648	-.0472192	.0758707
sharepoverty	3.321191	2.755484	1.21	0.228	-2.079459	8.721841
lessthanhs	-.009519	.018401	-0.52	0.605	-.0455842	.0265462
bachelorup	-.0323893	.0231745	-1.40	0.162	-.0778104	.0130319
_cons	-3.965883	1.974348	-2.01	0.045	-7.835534	-.0962317

The RMSE value equals 2540.67, which is a higher value than the base model (RMSE = 2456.76), suggesting that sorting the *shareold* variable into *share65to84up* and *share85up* is not the best approach in order to get the best fit for the Poisson regression line. The next step is to analyze the p-values of replaced coefficients.

Table 20: Chi-Square test for the joint effect of share65to84 and share85up on property2017

```
 ( 1)  [property2017]share65to842017 = 0
 ( 2)  [property2017]share85up = 0

           chi2(  2) =    2.10
         Prob > chi2 =    0.3506
```

Given that the p-value of the Chi-Square test for the joint effect of *share65to84* and *share 89up* equals to 0.3506 and that α = 0.05, there is not enough strong evidence to state that at least one of coefficients of these two variables should be different from zero. Furthermore, the RMSE value of this model is significantly higher than the baseline model, which supports the argument that replacing these the *shareold* variable would be a better approach to the analysis, this variable had a p-value of 0.162 on the baseline model which is 50% lower 0.3506 than the p-value of the joint effect and shows shareold is more significant to the analysis.

Table 21: Chi-Square test comparison of coefficients of variables share65to84 and share85up

```
 ( 1)  [property2017]share65to842017 - [property2017]share85up = 0

           chi2(  1) =    0.00
         Prob > chi2 =    0.9700
```

The comparison of *share65to84* and *share85up* supports the argument that these two variables should not replace shareold. Since the p-value of the Chi-Square test is 0.97, there is not enough substantial evidence to state that these two variables are different.

Table 22: Robust-Poisson regression replacing sharepoverty with sharepoor and shareverypoor

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(9)      =     892.94
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14967.701               Pseudo R2         =     0.9018
```

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.31094	.1466636	8.94	0.000	1.023484	1.598395
popgrowth	.3918336	15.79644	0.02	0.980	-30.56863	31.3523
shareyoung	-20.70741	11.62021	-1.78	0.075	-43.4826	2.067792
shareold	-4.151512	2.901284	-1.43	0.152	-9.837925	1.534902
unemployment	.0092709	.0332433	0.28	0.780	-.0558847	.0744266
sharepoor	5.975349	3.128632	1.91	0.056	-.1566562	12.10735
shareverypoor	.5807046	5.986128	0.10	0.923	-11.15189	12.3133
lessthanhs	-.0195076	.0256606	-0.76	0.447	-.0698015	.0307863
bachelorup	-.0331551	.0257554	-1.29	0.198	-.0836348	.0173246
_cons	-3.855985	2.0274	-1.90	0.057	-7.829616	.1176458

The RMSE value equals 2443.41, which is a lower value than the base model (RMSE = 2456.76), suggesting that sorting the *sharepoverty* variable into *sharepoor* and *shareverypoor* is a better approach in order to get the best fit for the Poisson regression line. The next step is to analyze the p-values of replaced coefficients.

Table 23: Chi-Square test for the joint effect of shareverypoor and sharepoor on property2017

```
 ( 1)  [property2017]sharepoor = 0
 ( 2)  [property2017]shareverypoor = 0

           chi2(  2) =    3.54
         Prob > chi2 =    0.1705
```

Given that the p-value of the Chi-Square test for the joint effect of *sharepoor* and *share verypoor* equals to 0.1705 and that α = 0.05, there is not enough strong evidence to state that at least one the coefficients of these two variables should be different from zero. It is essencial to point out, though, that the RMSE value of this model is lower than the baseline model, the difference is approximately 0.55%.

Table 24: Chi-Square test comparison of coefficients of variables shareverypoor and sharepoor

```
 ( 1)  [property2017]sharepoor - [property2017]shareverypoor = 0

           chi2(  1) =    0.70
         Prob > chi2 =    0.4033
```

The comparison of *sharepoor* and *shareverypoor* indicate that these two variables are equal. Since the p-value of the Chi-Square test is 0.4033, it can be concluded that there is not enough strong evidence to disprove that these two variables are equal. Lastly, it is inconclusive if *sharepoverty* should be replaced or not by *sharepoor* or *shareverypoor* because the RMSE value, for this specific case drops when the replacement happens, but the hypotheses tested were inconclusive.

Table 25: Robust-Poisson regression replacing bachelorup with bachelor and graduatedegree

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(9)      =     839.03
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14915.857               Pseudo R2         =     0.9021
```

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.344068	.162523	8.27	0.000	1.025529	1.662607
popgrowth	2.554159	15.28801	0.17	0.867	-27.40979	32.51811
shareyoung	-20.73579	11.08906	-1.87	0.061	-42.46995	.9983705
shareold	-4.325531	2.902743	-1.49	0.136	-10.0148	1.36374
unemployment	.0112243	.0329624	0.34	0.733	-.0533808	.0758295
sharepoverty	2.241477	3.957434	0.57	0.571	-5.51495	9.997905
lessthanhs	-.0110515	.0208173	-0.53	0.596	-.0518527	.0297498
bachelor	-.0641927	.060353	-1.06	0.288	-.1824823	.0540969
graduatedegree	.0006243	.0599576	0.01	0.992	-.1168904	.118139
_cons	-3.96875	1.93969	-2.05	0.041	-7.770473	-.1670274

The RMSE value equals 2672.69, which is a higher value than the base model (RMSE = 2456.76), suggesting that sorting the *bachelorup* variable into *bachelor* and *graduate* is not the best approach in order to get the best fit for the Poisson regression line. The next step is to analyze the p-values of replaced coefficients.

Table 26: Chi-Square test for the joint effect of bachelor and greaduatedegree on property2017

```
 ( 1)  [property2017]bachelor = 0
 ( 2)  [property2017]graduatedegree = 0

           chi2(  2) =    1.91
         Prob > chi2 =    0.3843
```

Given that the p-value of the Chi-Square test for the joint effect of *bachelor* and *graduatedegree* equals to 0.3843 and that $\alpha = 0.05$, there is not enough strong evidence to state that at least one of the coefficients is different from zero. Furthermore, although the RMSE value is 8.7% higher while compared to the baseline model.

Table 27: Chi-Square test comparison of coefficients of variables bachelor and graduatedegree

```
( 1)  [property2017]bachelor - [property2017]graduatedegree = 0

           chi2(  1) =    0.35
         Prob > chi2 =    0.5533
```

The comparison of *barchelor* and *sgraduatedegree* supports the argument that these two variables should not replace *graduateup*. Since the p-value of the Chi-Square test is 0.5533, it can be concluded that there is not enough strong evidence to state that these two variables are different. Indeed, since the previous analyzed evidence has shown that the replacement variable does not yield not statistically significant changes to the model and a worse RMSE value, the best approach is not replacing the *graduateup* variable.

Table 28: Robust-Poisson regression replacing lessthanhs with somecollege, associate, and highschool

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(10)     =    1141.62
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14720.347               Pseudo R2         =     0.9034
```

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.292286	.1386717	9.32	0.000	1.020494	1.564077
popgrowth	.1665319	14.54772	0.01	0.991	-28.34648	28.67955
shareyoung	-22.99325	12.54162	-1.83	0.067	-47.57437	1.58786
shareold	-4.09294	2.953005	-1.39	0.166	-9.880723	1.694842
unemployment	-.0061053	.0522392	-0.12	0.907	-.1084923	.0962817
sharepoverty	4.34683	3.668067	1.19	0.236	-2.842448	11.53611
highschool	-.0179484	.032222	-0.56	0.578	-.0811024	.0452056
somecollege	.0322355	.0439486	0.73	0.463	-.0539021	.1183731
associate	.026788	.0539285	0.50	0.619	-.07891	.132486
bachelorup	-.030935	.0246739	-1.25	0.210	-.0792948	.0174249
_cons	-4.013571	2.798726	-1.43	0.152	-9.498973	1.47183

The RMSE value equals 2644.23, which is a higher value than the base model (RMSE = 2456.76), suggesting that sorting the *lessthanhs* variable into *highschool, somecollege* and *associate* is not the best approach in order to get the best fit for the Poisson regression line. The next step is to analyze the p-values of replaced coefficients.

Table 29: Chi-Square test for the joint effect of somecollege, highschool, and associate on property2017

```
 ( 1)  [property2017]somecollege = 0
 ( 2)  [property2017]highschool = 0
 ( 3)  [property2017]associate = 0

           chi2(  3) =    0.84
         Prob > chi2 =    0.8404
```

Given that the p-value of the Chi-Square test for the joint effect of *somecollege, highschool,* and *associate* equals to 0.8404 and that α = 0.05, there is not enough strong evidence to state that at least one of the three coefficients is different from zero. Furthermore, although the RMSE value is 7.6% higher while compared to the baseline model, supporting the argument that these three variables should not replace *lessthanhs*.

Table 30: RMSE values for tested models

Model	Replaced Variable	RMSE	%Difference related to the baseline model
Model1	shareyoung	2446.43	-0.4%
Model2	shareold	2540.67	3.4%
Model3	sharepoverty	2443.41	-0.5%
Model4	bachelorup	2672.69	8.7%
Model5	lessthanhs	2644.23	7.6%
Baseline model	-	2456.76	0.0%

This RMSE table helps understanding the impact of each replaced variable on the predictive power of each model. The percentage difference column shows that the replacement variables for *lessthanhs* and *bachelorup* presented a highly-positive percentage difference when compared to the baseline model, so they were not considered to the analysis. The case of *shareold*'s replacement variables is similar since it yields a RMSE value 3.4% higher while compared to the base line model. Lastly, the considered variables for a possible replacement are *shareyoung* and *sharepoverty*, these presented a negative percentage difference what is a high indicator that these should increase the predictive power of a model.

3.5 Model Selection for Socioeconomic factors

Table 31: Socioeconomic models comparison - Output table for 4 different tested models

	(1) property2017	(2) property2017	(3) property2017	(4) property2017	(5) property2017
property2017					
lnpop2017	1.318***	1.032***	1.031***	1.031***	1.036***
	(0.151)	(0.129)	(0.128)	(0.129)	(0.089)
popgrowth	2.496	6.422			
	(14.503)	(12.687)			
shareyoung	-21.79	-18.42*	-18.69*	-18.76	
	(12.318)	(9.117)	(9.166)	(9.643)	
shareold	-4.245	-7.899	-8.483	-8.508	-8.420*
	(3.013)	(4.495)	(4.534)	(4.881)	(4.198)
unemployment	0.0144	0.00116			
	(0.036)	(0.033)			
sharepoverty	3.231				4.392*
	(3.535)				(1.840)
lessthanhs	-0.0102				
	(0.020)				
bachelorup	-0.0330				
	(0.026)				
sharepoor		8.193**	7.764**	7.637**	
		(3.120)	(2.879)	(2.899)	
shareverypoor		0.582	-0.230		
		(5.641)	(4.722)		
share14to17					-16.54
					(18.802)
share18to24					-20.71
					(11.088)
_cons	-3.900	-1.769	-1.464	-1.455	-1.496
	(2.023)	(2.950)	(2.804)	(2.942)	(2.376)
RMSE	2456.76	2302.10	2241.39	2083.31	2489.41
N	95	95	95	95	95

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

This regression table above provides the statistical significance of each coefficient and the RMSE value, which will help to analyze the statistical power of each of the Poisson regression models. The baseline model has a RMSE value of 2456.76, and the variables that were subject to change were *sharepoverty; i*n other models it was replaced by *sharepoor* and *shareverypoor*, and *shareyoung* which was replaced by *share14to17* and *share18to24*.

On first tested model (2), the *sharepoverty* variable was replaced and the *lessthanhs* and *graduateup* variable were removed from the model as both presented high p-values and also the absolute value of the Pearson's correlation r with the poverty variables were above 0.5 (respectively: r = 0.6386 and r = -0.6269), this fact indicates high degree of collinearity between variables, the *popgrowth* variable was also removed due to high degree of collinearity with *lnpop2017*(r =0.5614) and *povertyshare* (r = -0.6938). The RMSE value for this model is 2302.1, and it decreased 6.3% when compared to the baseline model.

Since the first test model presented a consistent result, the prediction power of it can be improved, so on the second tested model (3) the *unemployment* variable was removed as its p-value was equal to 0.970 indicating that there is strong evidence that this coefficient is statistically insignificant to the analysis. The RMSE value for this model is 2241.39, and it decreased 2.6% when compared to the model (2).

The last change that was made to this model involving the poverty data was the removal of variable *shareverypoor* as its p-value was 0.946 indicating that there is strong evidence that this coefficient is statistically insignificant to the analysis. The RMSE value for this model is 2083.31, it decreased 7.1% when compared to the model (3).

Lastly, on model (5) the *shareyoung* variable was replaced by *share14to17* and share18to24, and the *sharepoverty* variable was back transformed from the previous replacement it was applied to it. The RMSE value for this Poisson model was 2489.41, which is higher than the base model, this fact shows that the replacement of the *shareyoung* variable will not be useful for the prediction, but the replacement of *sharepoverty* will.

The chosen model involving socio-economic data is the Robust-Poisson regression of property2017 on *shareyoung, shareold, lnpop2017* and *sharepoor* (Table 29). This model presented the best RMSE values, and the coefficients are statistically significant to the analysis as their p-values are all below 0.052.

Table 32: Regression of shareyoung, shareold, lnpop2017, and sharepoor on property2017

```
Poisson regression                              Number of obs     =         96
                                                Wald chi2(4)      =     224.03
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -16058.109               Pseudo R2         =     0.8946
```

property2017	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
lnpop2017	1.032805	.1288242	8.02	0.000	.7803143	1.285296
shareyoung	-18.71198	9.628945	-1.94	0.052	-37.58436	.1604101
shareold	-8.499526	4.889626	-1.74	0.082	-18.08302	1.083964
sharepoor	7.658722	2.89849	2.64	0.008	1.977787	13.33966
_cons	-1.484192	2.9412	-0.50	0.614	-7.248837	4.280454

3.4 Overall model selection

While combining the strongest social, demographic, and economic data, it was clear that some specific variables such as *sharepoor*, *shareyoung*, and *shareold* and verypoor have high degree of correlation with property crimes and help on the prediction models.

The table above compares three different models. The second model, which seems a good option to be selected since its Pseudo-R2 and the RMSE values are the best while compared to the other models; the RMSE indicates less variance on the line and a more precise prediction. Furthermore, this model accounts for important variables highly correlated to property crime rates. Lastly, this model specifies the different variable that has a direct impact on the crime rate, and it can also be noticed by the significance of the coefficients:

Table 33: Correlation matrix for 9 different socio-economic variables

	pro~2018	proper~h	lnprop~7	sharep~r	sharev~r	sharey~g	unempl~t
property2018	1.0000						
propertygrowth	0.0382	1.0000					
lnproperty2017	0.6455	0.1144	1.0000				
sharepoor	-0.1873	0.0638	-0.4072	1.0000			
shareverypoor	-0.1809	0.1448	-0.3593	0.5528	1.0000		
shareyoung	0.0616	-0.0588	0.1910	-0.0767	0.1189	1.0000	
unemployment	-0.0739	0.2784	-0.1644	0.3105	0.4388	0.0863	1.0000

From the correlation matrix above, it possible to gather some insight based on the Person's r for the relation of each of the coefficients. The *lnproperty2017* variable is highly correlated with *property2018* (r = 0.6455) as both refer to property crimes' data, but for different years. Lastly, sharepoor is highly correlated with *shareverypoor* (r = 0.5528), and this happened because both refer to poverty a variable.

The correlation matrix of the baseline models for both historical factors and socioeconomic presented multiples cases of high correlated variables. After multiples tests and model selections, the cases of collinearity among the model's variables decreased significantly. The last step is the comparison of the combined models using the most significant variables that were listed and compared above.

Table 34: Combined models comparison - Output table for 4 different tested models

	(1) property2018	(2) property2018	(3) property2018	(4) property2018
property2018				
propertygrowth	-0.505	-0.493	-0.358	-0.499
	(0.494)	(0.474)	(0.427)	(0.385)
lnproperty2017	0.998***	0.998***	0.995***	0.996***
	(0.010)	(0.010)	(0.010)	(0.011)
sharepoverty		0.0233		
		(0.315)		
sharepoor			2.133	2.307
			(1.496)	(1.451)
shareverypoor			-2.342	-2.636
			(1.747)	(1.680)
shareyoung				1.287
				(0.795)
_cons	-0.118	-0.123	-0.120	-0.310*
	(0.076)	(0.103)	(0.091)	(0.128)
RMSE	468.81	475.76	449.35	505.80
N	95	95	95	95

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The tables above show the comparison of loocv RMSE values of the generated combined models. Model (1) is the best fit for the historic of property crime's factors, it contains only *propertygrowth* and *lnproperty2017* as factors, this model contained the best RMSE value from all previously tested, so that it will be adopted as the baseline model.

Model (2) included the *sharepoverty* variable since poverty is a factor that helps explain why property crimes happen, but the RMSE value went from 468.81 to 475.76, which shows that this model is better than the baseline. Since it was tested before that the *sharepoverty* variable works better to the model when it is replaced by *sharepoor* and *shareverypoor*, this approach was selected for the model (3). This was the model for the analysis, it presented the best RMSE value, 449.35, meaning this one is the model with high predictive power.

Lastly, when the *shareyoung* variable was introduced on model (4) to check if it would increase the prediction power of the model, it was possible to conclude that it would not benefit the predictions since the RMSE went up to 505.80. Finally, model (3) is the selected model as it has the highest RMSE value, and its coefficients are significant to the analysis when joint tested.

4. Discussion

The values of standard deviation, mean, and RMSE values were inflated due to outliers on the dataset, which were the highest populated counties within Georgia. Since the objective of this study is to predict the property crime rates within all counties in Georgia, these highly-populated counties were maintained in the dataset.

After extensive research and testing of models, the best models for predicting the rate of property crimes in each county within Georgia were determined. The most telling factors, which used the historic predictors for the model, were the natural log of property crimes for 2017 and the growth of property crimes between 2012 and 2017. Apart from the combined model, which included both historic and socioeconomic factors, this was the best-fit among all others.

Finally, in the case of the most telling predictors for the combined model, it took into consideration the best socioeconomic predictor, the ones that reduced the RMSE of the tested model when compared to the baseline. It incorporated the best telling predictors with the historic predictors' model, aiming at finding the best fit model with the highest prediction power. It was explicitly found specifically that the share of people under half of the poverty's level income and between half and the poverty's income aggregated with the historic predictors provided the best fit and highest prediction power.

5. Conclusion

Overall, a study was conducted for Georgia's state government in order to come up with the best model in order to predict the property crime and investigate which are the factor that impacted by increasing the property crime rates, so it is possible to allocate resources to solve the problem at its source. Furthermore, the socioeconomic factor that had the greatest impact on prediction was the share of people under half of the poverty's level income, and between half and the poverty's income aggregated with the historic predictors provided the best fit and highest prediction power. Finally, it can conclude that an efficient way to decrease the property crime rates in county within Georgia, it is reducing the poverty levels.

Appendix A: Property Crimes Predictions for 2019 grouped by county

county	property2019
Baldwin	493.4015
Banks	348.1512
Ben Hill	161.0253
Berrien	194.099
Bleckley	91.93798
Brooks	125.8582
Bryan	289.0164
Bulloch	438.6293
Butts	295.22
Charlton	102.8558
Chattooga	87.65781
Cherokee	1178.466
Clay	10.84623
Clayton	7487.975
Clinch	39.69954
Cobb	8294.769
Coffee	438.3635
Columbia	1593.814
Cook	119.1769
Coweta	910.0548
Crawford	182.8411
Crisp	396.7691
Dawson	396.3569
Decatur	193.1077
DeKalb	19190.68
Dodge	173.9161
Dooly	81.43768
Dougherty	365.8154
Early	94.24102
Echols	46.51483
Effingham	303.3893
Elbert	256.9738
Emanuel	210.207
Evans	31.36139
Fannin	186.2782
Fayette	445.3214
Fulton	1386.254
Gilmer	388.2971
Glynn	1509.402

Grady	166.7265
Greene	106.3729
Gwinnett	12750.62
Habersham	309.1581
Hall	1453.875
Hancock	91.71284
Henry	3141.025
Houston	834.4865
Irwin	92.79786
Jasper	155.1048
Jeff Davis	94.94231
Jefferson	77.11692
Jones	227.8505
Lamar	100.5915
Lanier	127.3118
Laurens	450.9486
Lee	467.8004
Liberty	205.5602
Long	145.3012
Madison	362.5175
Mitchell	231.5638
Monroe	279.5061
Newton	1385.935
Oconee	398.8565
Oglethorpe	173.4772
Paulding	1777.885
Peach	204.4712
Pickens	365.9712
Pierce	236.7844
Pike	177.6722
Polk	504.1217
Pulaski	197.6674
Rabun	191.8175
Rockdale	1132.404
Schley	9.744584
Seminole	53.10657
Spalding	759.3322
Stephens	243.8181
Sumter	168.2458
Talbot	42.39161
Taylor	42.55139
Thomas	427.7527
Towns	123.145
Treutlen	40.95578

Troup	567.9871
Twiggs	93.70404
Union	195.2697
Upson	260.6571
Walton	518.225
Ware	528.0677
Warren	27.97522
Wayne	631.7867
Webster	1.2239
White	123.7384
Whitfield	1359.502
Wilkes	19.2395
Wilkinson	34.35394

Appendix B: Do-file-for-Project

```
/* QMB 3200 Regression Example Work */

clear

cd "C:\Users\luizg\Desktop\QMB Project"

import delimited "C:\Users\luizg\Desktop\QMB Project\projectdata.csv",
encoding(ISO-8859-2)

*Generating Ln Variables and Variable growth

gen lnproperty2012 = ln(property2012)
gen lnproperty2017 = ln(property2017)
gen lnviolent2012 = ln(violent2012)
gen lnviolent2017 = ln(violent2017)
gen propertygrowth=(lnproperty2017-lnproperty2012)/5
gen violentgrowth=(lnviolent2017-lnviolent2012)/5
gen lnpop2012 = ln(pop2012)
gen lnpop2017 = ln(pop2017)
gen popgrowth=(lnpop2017-lnpop2012)/5

*Variables Summary

sum violent2012 violent2017 violent2018 violentgrowth lnviolent2012
lnviolent2017 popgrowth propertygrowth lnpop2012 lnpop2017
lnproperty2012 lnproperty2017 property2012 property2017 property2018
unemployment poor verypoor poppov graduatedegree bachelor associate
somecollege highschool edu9to12th eduless9th pop2012 pop2017
pop14to172012 pop14to172017 pop18to242012 pop18to242017 pop65up2017
pop65up2012 pop85up2012 pop85up2017
```

```
*Variables Description

describe violent2012 violent2017 violent2018 violentgrowth
lnviolent2012 lnviolent2017 popgrowth propertygrowth lnpop2012
lnpop2017 lnproperty2012 lnproperty2017 property2012 property2017
property2018 unemployment poor verypoor poppov graduatedegree bachelor
associate somecollege highschool edu9to12th eduless9th pop2012 pop2017
pop14to172012 pop14to172017 pop18to242012 pop18to242017 pop65up2017
pop65up2012 pop85up2012 pop85up2017

*Generating Table 4 Variables:

gen share14to17=pop14to172017/pop2017
gen share18to24=pop18to242017/pop2017
gen pop65to842017 = pop65up2017 - pop85up2017
gen share65to84=pop65to842017/pop2017
gen share85up=pop85up2017/pop2017
gen shareyoung = share14to17+share18to24
gen shareold = share65to84+share85up
gen shareverypoor = verypoor/poppov
gen sharepoor = poor/poppov
gen sharepoverty = sharepoor + shareverypoor
gen bachelorup = bachelor + graduate
gen lessthanhs = 100-highschool-somecollege-associate-bachelor-
graduate

sum pop65to842017 share14to17 share18to24 share65to84 share85up
shareold sharepoor sharepoverty shareverypoor shareyoung bachelorup
lessthanhs

* 3.1.1 Initial Models:

correlate property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth lnpop2017 popgrowth
```

```
poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth lnpop2017 popgrowth, robust

test lnproperty2017 propertygrowth

test lnviolent2017 violentgrowth

loocv poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth lnpop2017 popgrowth, robust

estimates store modelt1

poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth, robust

test lnviolent2017 violentgrowth

loocv poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth, robust

estimates store modelt2

poisson property2018 propertygrowth lnproperty2017, robust

loocv poisson property2018 propertygrowth lnproperty2017, robust

estimates store modelt3

esttab modelt* using "modelt.ttf" , se(3) replace

correlate property2018 propertygrowth lnproperty2017


* 3.2.2 Multicollinearity on the models


correlate property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup

reg property2017 lnpop2017 popgrowth shareyoung shareold unemployment
sharepoverty lessthanhs bachelorup

vif


*3.2.3 Baseline Model

poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust

estimates store modelr1
```

```
*3.2.4 Tests of alternatives for each class of independent variable

*Share young

poisson property2017 lnpop2017 popgrowth share14to17 share18to24
shareold unemployment sharepoverty lessthanhs bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth share14to17 share18to24
shareold unemployment sharepoverty lessthanhs bachelorup, robust

estimates store modelr2

test share18to24 share14to17

test share18to24 == share14to17


*Share old

poisson property2017 lnpop2017 popgrowth shareyoung share65to84
share85up unemployment sharepoverty lessthanhs bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung share65to84
share85up unemployment sharepoverty lessthanhs bachelorup, robust

estimates store modelr3

test share65to84 share85up

test share65to84 == share85up


*Share Poverty

poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor lessthanhs bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor lessthanhs bachelorup, robust

estimates store modelr4

test sharepoor shareverypoor

test sharepoor == shareverypoor


*Bachelorup

poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelor graduate, robust
```

```
loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty lessthanhs bachelor graduate, robust

estimates store modelr5

test bachelor graduate

test bachelor == graduate


*LessthanHS

poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty highschool somecollege associate bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty highschool somecollege associate bachelorup, robust

estimate store modelr6

test somecollege highschool associate


esttab modelr* using "modelr.ttf" , se(3) replace


*3.2.5 Estimate the best model

poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty lessthanhs bachelorup, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty lessthanhs bachelorup, robust

estimates store modelz1


poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoor shareverypoor, robust

loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoor shareverypoor, robust

estimate store modelz2


poisson property2017 lnpop2017 shareyoung shareold sharepoor shareverypoor, robust
```

```
loocv poisson property2017 lnpop2017 shareyoung shareold sharepoor
shareverypoor, robust

estimate store modelz3


poisson property2017 lnpop2017 shareyoung shareold sharepoor, robust

loocv poisson property2017 lnpop2017 shareyoung shareold sharepoor,
robust

estimate store modelz4


poisson property2017 lnpop2017 share14to17 share18to24 shareold
sharepoverty, robust

loocv poisson property2017 lnpop2017 share14to17 share18to24 shareold
sharepoverty, robust

estimate store modelz5

esttab modelz* using "modelz.rtf" , se(3) replace


*3.3 Prediction using past crime and other correlates


poisson property2018 propertygrowth lnproperty2017, robust

loocv poisson property2018 propertygrowth lnproperty2017, robust

estimates store modelt1

estat ic


poisson property2018 propertygrowth lnproperty2017 sharepoverty,
robust

loocv poisson property2018 propertygrowth lnproperty2017 sharepoverty,
robust

estimates store modelt2

estat ic


poisson property2018 propertygrowth lnproperty2017 sharepoor
shareverypoor, robust
```

```
loocv poisson property2018 propertygrowth lnproperty2017 sharepoor shareverypoor, robust

estimates store modelt3

estat ic


poisson property2018 propertygrowth lnproperty2017 sharepoor shareverypoor shareyoung, robust

loocv poisson property2018 propertygrowth lnproperty2017 sharepoor shareverypoor shareyoung, robust

estimates store modelt4

estat ic

esttab modelt* using "modelt.rtf" , se(3) replace


*3.4 Predictions for 2018

gen lnproperty2018 = ln(property2018)

gen CrimeRate = (lnproperty2018-lnproperty2012)/6


gen property2019=exp(-0.364423*propertygrowth+0.9952104*lnproperty2018+2.1292*sharepoor-2.338646*shareverypoor-0.1199882)


list county property2019 (3), compress


STOP

log close
```

Appendix C: Log-file-for-Project

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   16.0   Copyright 1985-2019 StataCorp
LLC
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC
http://www.stata.com
                                      979-696-4600
stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata license expires 16 Mar 2020:
       Serial number:  301609236389
         Licensed to:  Luiz Gustavo Fagundes Malpele
                       Florida Polytechnic University

Notes:
      1.  Unicode is supported; see help unicode_advice.

. doedit "C:\Users\luizg\Desktop\QMB Project\do-file-project.do"

. do "C:\Users\luizg\AppData\Local\Temp\STD5764_000000.tmp"

. /* QMB 3200 Regression Example Work */
.
. clear
```

```
.
. cd "C:\Users\luizg\Desktop\QMB Project"
C:\Users\luizg\Desktop\QMB Project


.
. import delimited "C:\Users\luizg\Desktop\QMB
Project\projectdata.csv", encoding(ISO-8859-2)
(28 vars, 96 obs)


.
. *Generating Ln Variables and Variable growth
.
. gen lnproperty2012 = ln(property2012)


. gen lnproperty2017 = ln(property2017)


. gen lnviolent2012 = ln(violent2012)
(2 missing values generated)


. gen lnviolent2017 = ln(violent2017)
(2 missing values generated)


. gen propertygrowth=(lnproperty2017-lnproperty2012)/5


. gen violentgrowth=(lnviolent2017-lnviolent2012)/5
(4 missing values generated)


. gen lnpop2012 = ln(pop2012)


. gen lnpop2017 = ln(pop2017)
```

```
. gen popgrowth=(lnpop2017-lnpop2012)/5


.

. *Variables Summary

.

. sum violent2012 violent2017 violent2018 violentgrowth lnviolent2012
lnviolent2017 popgrowth propertygrowth lnpop2012 lnpop2017
lnproperty2012 lnproperty2017 property2012 property2017 p

> roperty2018 unemployment poor verypoor poppov graduatedegree
bachelor associate somecollege highschool edu9to12th eduless9th
pop2012 pop2017 pop14to172012 pop14to172017 pop18to242012 p

> op18to242017 pop65up2017 pop65up2012 pop85up2012 pop85up2017


    Variable |        Obs        Mean    Std. Dev.       Min
Max

-------------+---------------------------------------------------------
-

 violent2012 |         96    119.8021     402.174          0
3420

 violent2017 |         96    122.5208    388.4895          0
3184

 violent2018 |         96    128.8438    389.6356          1
3085

violentgro~h |         92    .0291374    .1216718  -.3531568
.2834132

lnviole~2012 |         94    3.489516    1.394921          0
8.137396

-------------+---------------------------------------------------------
-

lnviole~2017 |         94    3.626938    1.279686   .6931472
8.065893

   popgrowth |         96    .0033757    .0107121  -.0135521
.0322412

propertygr~h |         96   -.0349247    .0944389  -.5278115
.5051457
```

```
   lnpop2012 |         96    10.28049    1.192467   7.925158
13.78838
   lnpop2017 |         96    10.29737    1.221712   7.861727
13.85366
-------------+---------------------------------------------------------
-
lnprope~2012 |         96    5.888458    1.467278   .6931472
10.31841
lnprope~2017 |         96    5.713834    1.444298   1.098612
10.05389
property2012 |         96    1271.885    3672.148          2
30285
property2017 |         96    1109.042    3161.136          3
23246
property2018 |         96         978    2877.916          1
21426
-------------+---------------------------------------------------------
-
unemployment |         96     8.20625    3.004481          4
20.3
        poor |         96     6569.26    13266.33        322
79132
    verypoor |         96     5385.75    11167.23        187
77387
      poppov |         96    74895.93    167551.8       2613
978533
graduatede~e |         96    7.464583    3.789653        2.1
25.2
-------------+---------------------------------------------------------
-
    bachelor |         96    11.43646    5.627566          3
29.9
   associate |         96    7.396875    1.521388        3.4
12.2
 somecollege |         96    20.58437    3.261234         12
31.1
```

```
  highschool |         96    35.37396     6.50032       18.3
52.1
  edu9to12th |         96    11.85208    3.774289        3.2
22.8
-------------+---------------------------------------------------------
-
  eduless9th |         96    5.896875    2.476859        1.5
14
     pop2012 |         96    74038.13    163019.8       2766
973236
     pop2017 |         96     78400.6    175426.6       2596
1038884
pop14~172012 |         96    4176.781    9046.742        119
54311
pop14~172017 |         96    4445.188    9903.134        149
60463
-------------+---------------------------------------------------------
-
pop18~242012 |         96    7235.083    16102.45        216
105314
pop18~242017 |         96    7289.813    16524.48        190
105701
 pop65up2017 |         96    10273.48    19364.88        493
118228
 pop65up2012 |         96     8192.01     14809.5        412
93402
 pop85up2012 |         96    881.0833    1755.629         33
12704
-------------+---------------------------------------------------------
-
 pop85up2017 |         96    1028.354      2052.7         46
14101


.
. *Variables Description
.
```

```
. describe violent2012 violent2017 violent2018 violentgrowth
lnviolent2012 lnviolent2017 popgrowth propertygrowth lnpop2012
lnpop2017 lnproperty2012 lnproperty2017 property2012 property2

> 017 property2018 unemployment poor verypoor poppov graduatedegree
bachelor associate somecollege highschool edu9to12th eduless9th
pop2012 pop2017 pop14to172012 pop14to172017 pop18to242

> 012 pop18to242017 pop65up2017 pop65up2012 pop85up2012 pop85up2017


              storage   display    value
variable name   type    format     label      variable label

----------------------------------------------------------------------
----------------------------------------------------------------------
----------------------------------------------

violent2012     int     %8.0g                 Violent2012
violent2017     int     %8.0g                 Violent2017
violent2018     int     %8.0g                 Violent2018
violentgrowth   float   %9.0g
lnviolent2012   float   %9.0g
lnviolent2017   float   %9.0g
popgrowth       float   %9.0g
propertygrowth  float   %9.0g
lnpop2012       float   %9.0g
lnpop2017       float   %9.0g
lnproperty2012  float   %9.0g
lnproperty2017  float   %9.0g
property2012    int     %8.0g                 Property2012
property2017    int     %8.0g                 Property2017
property2018    int     %8.0g                 Property2018
unemployment    float   %9.0g                 Unemployment
poor            long    %12.0g                Poor
verypoor        long    %12.0g                VeryPoor
poppov          long    %12.0g                PopPov
```

```
graduatedegree  float   %9.0g                 Graduate degree
bachelor        float   %9.0g                 Bachelor
associate       float   %9.0g                 Associate
somecollege     float   %9.0g                 SomeCollege
highschool      float   %9.0g                 HighSchool
edu9to12th      float   %9.0g                 Edu9to12th
eduless9th      float   %9.0g                 EduLess9th
pop2012         long    %12.0g                Pop2012
pop2017         long    %12.0g                Pop2017
pop14to172012   long    %12.0g                Pop14to172012
pop14to172017   long    %12.0g                Pop14to172017
pop18to242012   long    %12.0g                Pop18to242012
pop18to242017   long    %12.0g                Pop18to242017
pop65up2017     long    %12.0g                Pop65up2017
pop65up2012     long    %12.0g                Pop65up2012
pop85up2012     int     %8.0g                 Pop85up2012
pop85up2017     int     %8.0g                 Pop85up2017

.
. *Generating Table 4 Variables:
.
. gen share14to17=pop14to172017/pop2017

. gen share18to24=pop18to242017/pop2017

. gen pop65to842017 = pop65up2017 - pop85up2017

. gen share65to84=pop65to842017/pop2017

. gen share85up=pop85up2017/pop2017
```

```
. gen shareyoung = share14to17+share18to24

. gen shareold = share65to84+share85up

. gen shareverypoor = verypoor/poppov

. gen sharepoor = poor/poppov

. gen sharepoverty = sharepoor + shareverypoor

. gen bachelorup = bachelor + graduate

. gen lessthanhs = 100-highschool-somecollege-associate-bachelor-
graduate


.
. sum pop65to842017 share14to17 share18to24 share65to84 share85up
shareold sharepoor sharepoverty shareverypoor shareyoung bachelorup
lessthanhs


    Variable |        Obs        Mean    Std. Dev.       Min
Max
-------------+---------------------------------------------------------
-
pop65~842017 |         96    9245.125     17337.7        447
104127
 share14to17 |         96    .0540882    .0076406   .0325851
.07153
 share18to24 |         96    .0883316    .0250651   .0592906
.2374612
 share65to84 |         96    .1551222    .0426909   .0784637
.306959
```

```
   share85up |         96    .0177731    .0063023   .0063191
.0387641
-------------+---------------------------------------------------------
-
    shareold |         96    .1728953      .04791   .0880881
.3445706
   sharepoor |         96    .1133421    .0386926    .037361
.2119249
sharepoverty |         96    .2069274    .0694555   .0630627
.4144205
shareveryp~r |         96    .0935853    .0401298   .0224568
.2611427
  shareyoung |         96    .1424199    .0248166    .100813
.2803217
-------------+---------------------------------------------------------
-
  bachelorup |         96    18.90104    9.120185        6.9
50.2
  lessthanhs |         96    17.74375    5.759857   5.000001
31.4


.
. * 3.1.1 Initial Models:
.
. correlate property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth lnpop2017 popgrowth
(obs=92)


             | pro~2018 lnv~2017 violen~h lnprop~7 proper~h lnpop2~7
popgro~h
-------------+---------------------------------------------------------
-------
property2018 |   1.0000
lnviole~2017 |   0.5904   1.0000
violentgro~h |  -0.0877   0.0578   1.0000
```

```
lnprope~2017 |   0.6762   0.7736  -0.2816   1.0000
propertygr~h |   0.0271  -0.0321   0.3266  -0.0654   1.0000
   lnpop2017 |   0.6171   0.7295  -0.1841   0.8802  -0.1026   1.0000
   popgrowth |   0.2623   0.3206  -0.1024   0.4586  -0.0202   0.5404
1.0000


. poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth lnpop2017 popgrowth, robust


Iteration 0:   log pseudolikelihood = -42716.044
Iteration 1:   log pseudolikelihood = -8861.0855
Iteration 2:   log pseudolikelihood = -2235.6387
Iteration 3:   log pseudolikelihood = -2111.9449
Iteration 4:   log pseudolikelihood = -2111.8536
Iteration 5:   log pseudolikelihood = -2111.8536

Poisson regression                              Number of obs     =
92
                                                Wald chi2(6)      =
15617.55
                                                Prob > chi2       =
0.0000
Log pseudolikelihood = -2111.8536               Pseudo R2         =
0.9841

----------------------------------------------------------------------
----------
               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+------------------------------------------------------
----------
```

```
 lnviolent2017 |  -.0059011   .0109741    -0.54   0.591    -.0274099
.0156077
 violentgrowth |    .164006   .2144143     0.76   0.444    -.2562382
.5842503
lnproperty2017 |   1.106563   .0980341    11.29   0.000     .9144202
1.298707
propertygrowth |  -.3762712   .6124239    -0.61   0.539      -1.5766
.8240576
     lnpop2017 |  -.1453132   .1426851    -1.02   0.308    -.4249709
.1343446
     popgrowth |   7.329432   5.282404     1.39   0.165    -3.023889
17.68275
         _cons |   .7267536   .8546479     0.85   0.395    -.9483255
2.401833
----------------------------------------------------------------------
----------

. test lnproperty2017 propertygrowth

 ( 1)  [property2018]lnproperty2017 = 0
 ( 2)  [property2018]propertygrowth = 0

           chi2(  2) =  192.46
         Prob > chi2 =    0.0000

. test lnviolent2017 violentgrowth

 ( 1)  [property2018]lnviolent2017 = 0
 ( 2)  [property2018]violentgrowth = 0

           chi2(  2) =    0.59
         Prob > chi2 =    0.7455
```

```
. loocv poisson property2018 lnviolent2017 violentgrowth
lnproperty2017 propertygrowth lnpop2017 popgrowth, robust



 Leave-One-Out Cross-Validation Results

-----------------------------------------

        Method           |    Value

-------------------------+---------------

Root Mean Squared Errors |   697.87929

Mean Absolute Errors     |   204.75109

Pseudo-R2                |   .94305856

-----------------------------------------


. estimates store modelt1


. poisson property2018 lnviolent2017 violentgrowth lnproperty2017
propertygrowth, robust


Iteration 0:   log pseudolikelihood =  -38489.32

Iteration 1:   log pseudolikelihood =  -7171.226

Iteration 2:   log pseudolikelihood = -2385.1893

Iteration 3:   log pseudolikelihood = -2332.7717

Iteration 4:   log pseudolikelihood = -2332.6422

Iteration 5:   log pseudolikelihood = -2332.6422


Poisson regression                               Number of obs     =
92

                                                 Wald chi2(4)      =
11066.89

                                                 Prob > chi2       =
0.0000
```

```
Log pseudolikelihood = -2332.6422              Pseudo R2         =
0.9824


--------------------------------------------------------------------------------
----------
               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+----------------------------------------------------------------
----------
 lnviolent2017 |   -.022071   .0138086    -1.60   0.110    -.0491354
.0049933
 violentgrowth |   .0614741   .3090819     0.20   0.842    -.5443154
.6672635
lnproperty2017 |   1.019412   .0116899    87.20   0.000     .9964999
1.042324
propertygrowth |  -.5839831   .5853319    -1.00   0.318    -1.731213
.5632464
         _cons |  -.1763327   .0845377    -2.09   0.037    -.3420235
-.0106419
--------------------------------------------------------------------------------
----------


. test lnviolent2017 violentgrowth


 ( 1)  [property2018]lnviolent2017 = 0
 ( 2)  [property2018]violentgrowth = 0


           chi2(  2) =    4.44
         Prob > chi2 =    0.1089


. loocv poisson property2018 lnviolent2017 violentgrowth
lnproperty2017 propertygrowth, robust
```

```
 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   754.37242
Mean Absolute Errors     |   201.41258
Pseudo-R2                |   .93329208
-----------------------------------------

. estimates store modelt2

. poisson property2018 propertygrowth lnproperty2017, robust

Iteration 0:   log pseudolikelihood = -5557.0807
Iteration 1:   log pseudolikelihood = -2421.9652
Iteration 2:   log pseudolikelihood = -2398.3012
Iteration 3:   log pseudolikelihood =  -2398.292
Iteration 4:   log pseudolikelihood =  -2398.292

Poisson regression                                Number of obs     =
96
                                                  Wald chi2(2)      =
11022.00
                                                  Prob > chi2       =
0.0000
Log pseudolikelihood =  -2398.292                 Pseudo R2         =
0.9823

----------------------------------------------------------------------
----------
               |               Robust
```

```
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+------------------------------------------------------
----------
propertygrowth |   -.504605   .4927611    -1.02   0.306    -1.470399
.461189
lnproperty2017 |    .997609   .0103021    96.84   0.000     .9774172
1.017801
         _cons |  -.1176997   .0754387    -1.56   0.119    -.2655569
.0301574
----------------------------------------------------------------------
----------

. loocv poisson property2018 propertygrowth lnproperty2017, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   468.80571
Mean Absolute Errors     |   134.02932
Pseudo-R2                |   .97684202
-----------------------------------------

. estimates store modelt3

. esttab modelt* using "modelt.ttf" , se(3) replace
(output written to modelt.ttf)

. correlate property2018 propertygrowth lnproperty2017
(obs=96)
```

```
             | pro~2018 proper~h lnprop~7
-------------+---------------------------
property2018 |   1.0000
propertygr~h |   0.0382   1.0000
lnprope~2017 |   0.6455   0.1144   1.0000


.
. * 3.2.2 Multicollinearity on the models
.
. correlate property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup
(obs=96)


             | pro~2017 lnpop2~7 popgro~h sharey~g shareold unempl~t
sharep~y lessth~s bachel~p
-------------+-------------------------------------------------------------------
-------------------------
property2017 |   1.0000
   lnpop2017 |   0.6383   1.0000
   popgrowth |   0.2753   0.5614   1.0000
  shareyoung |   0.0622   0.3219   0.0777   1.0000
    shareold |  -0.3345  -0.5563  -0.2526  -0.4963   1.0000
unemployment |  -0.0782  -0.1848  -0.2512   0.0863  -0.0468   1.0000
sharepoverty |  -0.2184  -0.5184  -0.6938   0.0260   0.1813   0.4265
1.0000
  lessthanhs |  -0.2814  -0.5495  -0.6068  -0.2463   0.2130   0.0719
0.6386   1.0000
  bachelorup |   0.5103   0.7198   0.6480   0.2251  -0.2502  -0.2580
-0.6269  -0.7673   1.0000
```

```
. reg property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup

      Source |       SS           df       MS      Number of obs   =
96
-------------+----------------------------------   F(8, 87)        =
11.04
       Model |   478275551         8  59784443.9   Prob > F        =
0.0000
    Residual |   471038637        87   5414237.2   R-squared       =
0.5038
-------------+----------------------------------   Adj R-squared   =
0.4582
       Total |   949314188        95  9992780.92   Root MSE        =
2326.9

------------------------------------------------------------------------------
--------
property2017 |      Coef.   Std. Err.      t    P>|t|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
--------
   lnpop2017 |   1581.666   344.9898     4.58   0.000     895.9613
2267.37
   popgrowth |  -32219.31   34057.26    -0.95   0.347    -99911.81
35473.18
  shareyoung |  -31539.32   11917.74    -2.65   0.010    -55227.13   -
7851.514
    shareold |  -7048.415   6929.557    -1.02   0.312    -20821.66
6724.83
unemployment |   30.99466   95.09476     0.33   0.745    -158.0165
220.0058
sharepoverty |   9578.919   5969.844     1.60   0.112    -2286.792
21444.63
```

```
  lessthanhs |   68.07154   75.25931     0.90   0.368    -81.51449
217.6576
  bachelorup |   140.2868   52.82418     2.66   0.009     35.29305
245.2806
       _cons |  -15454.63   4898.474    -3.15   0.002    -25190.88   -
5718.382
----------------------------------------------------------------------
--------

. vif

    Variable |       VIF       1/VIF
-------------+----------------------
  bachelorup |      4.07    0.245550
  lessthanhs |      3.30    0.303298
   lnpop2017 |      3.12    0.320822
sharepoverty |      3.02    0.331493
   popgrowth |      2.34    0.428202
    shareold |      1.93    0.517072
  shareyoung |      1.53    0.651540
unemployment |      1.43    0.698171
-------------+----------------------
    Mean VIF |      2.59

.
. *3.2.3 Baseline Model
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust

Iteration 0:   log pseudolikelihood = -143334.31
Iteration 1:   log pseudolikelihood =  -78086.38
Iteration 2:   log pseudolikelihood = -20284.874
```

```
Iteration 3:   log pseudolikelihood = -15549.804

Iteration 4:   log pseudolikelihood = -15079.904

Iteration 5:   log pseudolikelihood = -15076.717

Iteration 6:   log pseudolikelihood = -15076.717


Poisson regression                                Number of obs     =
96

                                                  Wald chi2(8)      =
768.49

                                                  Prob > chi2       =
0.0000

Log pseudolikelihood = -15076.717                 Pseudo R2         =
0.9011


---------------------------------------------------------------------------
--------

             |               Robust

property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

-------------+-------------------------------------------------------------
--------

   lnpop2017 |   1.318093   .1512191     8.72   0.000     1.021709
1.614477

   popgrowth |   3.042099   14.39551     0.21   0.833    -25.17258
31.25678

  shareyoung |  -21.69349   12.29053    -1.77   0.078    -45.78247
2.395501

    shareold |  -4.202415   3.007807    -1.40   0.162    -10.09761
1.692777

unemployment |   .0147603   .0356234     0.41   0.679    -.0550603
.0845809

sharepoverty |   3.259106   3.537914     0.92   0.357    -3.675078
10.19329

  lessthanhs |  -.0096915   .0197519    -0.49   0.624    -.0484045
.0290215
```

```
  bachelorup |  -.0328026   .0256176    -1.28   0.200    -.0830121
.017407
       _cons |  -3.954115   2.013414    -1.96   0.050    -7.900335   -
.0078954
----------------------------------------------------------------------
--------


. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust



 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2456.7627
Mean Absolute Errors     |   660.07151
Pseudo-R2                |   .47521604
-----------------------------------------


. estimates store modelr1


.
. *3.2.4 Tests of alternatives for each class of independent variable
. *Share young
. poisson property2017 lnpop2017 popgrowth share14to17 share18to24
shareold unemployment sharepoverty lessthanhs bachelorup, robust


Iteration 0:   log pseudolikelihood = -174276.51
Iteration 1:   log pseudolikelihood = -98162.909
Iteration 2:   log pseudolikelihood = -40675.962
Iteration 3:   log pseudolikelihood = -19008.199
```

```
Iteration 4:   log pseudolikelihood = -15197.961
Iteration 5:   log pseudolikelihood = -14688.787
Iteration 6:   log pseudolikelihood = -14678.961
Iteration 7:   log pseudolikelihood = -14678.947
Iteration 8:   log pseudolikelihood = -14678.947

Poisson regression                               Number of obs     =
96
                                                 Wald chi2(9)      =
653.97
                                                 Prob > chi2       =
0.0000
Log pseudolikelihood = -14678.947                Pseudo R2         =
0.9037

----------------------------------------------------------------------
--------
             |               Robust
property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+--------------------------------------------------------
--------
   lnpop2017 |   1.335356   .1607283     8.31   0.000     1.020334
1.650378
   popgrowth |   5.546575    14.1491     0.39   0.695    -22.18514
33.27829
  share14to17 |   -43.5058   20.71389    -2.10   0.036    -84.10429   -
2.907312
  share18to24 |  -18.51355   9.526645    -1.94   0.052    -37.18543
.158328
    shareold |   -5.65906   3.053151    -1.85   0.064    -11.64313
.3250053
unemployment |   .0031606   .0339404     0.09   0.926    -.0633614
.0696827
```

```
sharepoverty |   1.521039   3.434404     0.44   0.658    -5.210269
8.252347
  lessthanhs |  -.0097676   .0200316    -0.49   0.626    -.0490287
.0294936
  bachelorup |  -.0444096   .0269683    -1.65   0.100    -.0972666
.0084473
       _cons |  -2.326547   2.078942    -1.12   0.263    -6.401197
1.748104
----------------------------------------------------------------------
--------

. loocv poisson property2017 lnpop2017 popgrowth share14to17
share18to24 shareold unemployment sharepoverty lessthanhs bachelorup,
robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method           |    Value
-------------------------+---------------
Root Mean Squared Errors |   2446.4285
Mean Absolute Errors     |   686.0418
Pseudo-R2                |   .46033037
-----------------------------------------

. estimates store modelr2

. test share18to24 share14to17

 ( 1)  [property2017]share18to24 = 0
 ( 2)  [property2017]share14to17 = 0

           chi2(  2) =    5.88
```

```
         Prob > chi2 =    0.0528


. test share18to24 == share14to17


 ( 1)  - [property2017]share14to17 + [property2017]share18to24 = 0


           chi2(  1) =    1.68

         Prob > chi2 =    0.1947


.

. *Share old

. poisson property2017 lnpop2017 popgrowth shareyoung share65to84
share85up unemployment sharepoverty lessthanhs bachelorup, robust


Iteration 0:   log pseudolikelihood = -146695.75

Iteration 1:   log pseudolikelihood =   -92534.7

Iteration 2:   log pseudolikelihood = -20914.174

Iteration 3:   log pseudolikelihood = -15735.393

Iteration 4:   log pseudolikelihood = -15080.952

Iteration 5:   log pseudolikelihood = -15075.861

Iteration 6:   log pseudolikelihood =  -15075.86


Poisson regression                                Number of obs     =
96

                                                  Wald chi2(9)      =
1282.20

                                                  Prob > chi2       =
0.0000

Log pseudolikelihood =  -15075.86                 Pseudo R2         =
0.9011
```

```
------------------------------------------------------------------------------
             |               Robust
property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   lnpop2017 |   1.317202   .1421603     9.27   0.000     1.038573    1.595831
   popgrowth |   2.834533   17.93098     0.16   0.874    -32.30954     37.9786
  shareyoung |  -21.68312   12.20048    -1.78   0.076    -45.59563    2.229386
 share65to84 |  -3.912239    8.41269    -0.47   0.642    -20.40081    12.57633
   share85up |  -6.695603   74.94257    -0.09   0.929    -153.5803    140.1891
unemployment |   .0143257   .0314011     0.46   0.648    -.0472192    .0758707
sharepoverty |   3.321191   2.755484     1.21   0.228    -2.079459    8.721841
  lessthanhs |   -.009519    .018401    -0.52   0.605    -.0455842    .0265462
  bachelorup |  -.0323893   .0231745    -1.40   0.162    -.0778104    .0130319
       _cons |  -3.965883   1.974348    -2.01   0.045    -7.835534   -.0962317
------------------------------------------------------------------------------

. loocv poisson property2017 lnpop2017 popgrowth shareyoung share65to84 share85up unemployment sharepoverty lessthanhs bachelorup, robust

 Leave-One-Out Cross-Validation Results
```

```
-----------------------------------------
        Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2540.6655
Mean Absolute Errors     |   677.6374
Pseudo-R2                |   .45530684
-----------------------------------------

. estimates store modelr3

. test share65to84 share85up

 ( 1)  [property2017]share65to84 = 0
 ( 2)  [property2017]share85up = 0

           chi2(  2) =    2.10
         Prob > chi2 =    0.3506

. test share65to84 == share85up

 ( 1)  [property2017]share65to84 - [property2017]share85up = 0

           chi2(  1) =    0.00
         Prob > chi2 =    0.9700

.
. *Share Poverty
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor lessthanhs bachelorup, robust

Iteration 0:   log pseudolikelihood = -143828.02
```

```
Iteration 1:   log pseudolikelihood = -80706.745

Iteration 2:   log pseudolikelihood = -20900.264

Iteration 3:   log pseudolikelihood = -15642.056

Iteration 4:   log pseudolikelihood = -14970.321

Iteration 5:   log pseudolikelihood = -14967.701

Iteration 6:   log pseudolikelihood = -14967.701


Poisson regression                               Number of obs     =
96

                                                 Wald chi2(9)      =
892.94

                                                 Prob > chi2       =
0.0000

Log pseudolikelihood = -14967.701               Pseudo R2         =
0.9018


----------------------------------------------------------------------
---------

             |               Robust
 property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

--------------+-------------------------------------------------------
---------

    lnpop2017 |    1.31094   .1466636     8.94   0.000     1.023484
1.598395

    popgrowth |   .3918336   15.79644     0.02   0.980    -30.56863
31.3523

   shareyoung |  -20.70741   11.62021    -1.78   0.075     -43.4826
2.067792

     shareold |  -4.151512   2.901284    -1.43   0.152    -9.837925
1.534902

 unemployment |   .0092709   .0332433     0.28   0.780    -.0558847
.0744266

    sharepoor |   5.975349   3.128632     1.91   0.056    -.1566562
12.10735
```

```
shareverypoor |   .5807046   5.986128     0.10   0.923    -11.15189
12.3133
   lessthanhs |  -.0195076   .0256606    -0.76   0.447    -.0698015
.0307863
   bachelorup |  -.0331551   .0257554    -1.29   0.198    -.0836348
.0173246
        _cons |  -3.855985     2.0274    -1.90   0.057    -7.829616
.1176458
----------------------------------------------------------------------
---------


. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor lessthanhs bachelorup, robust



 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2443.4114
Mean Absolute Errors     |   660.71552
Pseudo-R2                |   .47995984
-----------------------------------------


. estimates store modelr4


. test sharepoor shareverypoor


 ( 1)  [property2017]sharepoor = 0
 ( 2)  [property2017]shareverypoor = 0


           chi2(  2) =    3.54
```

```
         Prob > chi2 =    0.1705


. test sharepoor == shareverypoor


 ( 1)  [property2017]sharepoor - [property2017]shareverypoor = 0


           chi2(  1) =    0.70
         Prob > chi2 =    0.4033


.
. *Bachelorup
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelor graduate, robust


Iteration 0:   log pseudolikelihood = -142105.03
Iteration 1:   log pseudolikelihood = -74916.708
Iteration 2:   log pseudolikelihood = -19151.781
Iteration 3:   log pseudolikelihood = -15375.271
Iteration 4:   log pseudolikelihood = -14920.499
Iteration 5:   log pseudolikelihood = -14915.859
Iteration 6:   log pseudolikelihood = -14915.857
Iteration 7:   log pseudolikelihood = -14915.857


Poisson regression                                Number of obs     =
96
                                                  Wald chi2(9)      =
839.03
                                                  Prob > chi2       =
0.0000
Log pseudolikelihood = -14915.857                 Pseudo R2         =
0.9021
```

```
------------------------------------------------------------------------------
               |               Robust
  property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------------+----------------------------------------------------------------
     lnpop2017 |   1.344068    .162523     8.27   0.000     1.025529    1.662607
     popgrowth |   2.554159   15.28801     0.17   0.867    -27.40979    32.51811
    shareyoung |  -20.73579   11.08906    -1.87   0.061    -42.46995    .9983705
      shareold |  -4.325531   2.902743    -1.49   0.136     -10.0148     1.36374
  unemployment |   .0112243   .0329624     0.34   0.733    -.0533808    .0758295
  sharepoverty |   2.241477   3.957434     0.57   0.571     -5.51495    9.997905
    lessthanhs |  -.0110515   .0208173    -0.53   0.596    -.0518527    .0297498
      bachelor |  -.0641927    .060353    -1.06   0.288    -.1824823    .0540969
graduatedegree |   .0006243   .0599576     0.01   0.992    -.1168904     .118139
         _cons |   -3.96875    1.93969    -2.05   0.041    -7.770473   -.1670274
------------------------------------------------------------------------------

. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold unemployment sharepoverty lessthanhs bachelor graduate, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
```

```
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2672.6932
Mean Absolute Errors     |   719.37666
Pseudo-R2                |   .3819513
-----------------------------------------

. estimates store modelr5

. test bachelor graduate

 ( 1)  [property2017]bachelor = 0
 ( 2)  [property2017]graduatedegree = 0

           chi2(  2) =    1.91
         Prob > chi2 =    0.3843

. test bachelor == graduate

 ( 1)  [property2017]bachelor - [property2017]graduatedegree = 0

           chi2(  1) =    0.35
         Prob > chi2 =    0.5533

.
. *LessthanHS
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty highschool somecollege associate bachelorup,
robust

Iteration 0:   log pseudolikelihood = -153878.96
```

```
Iteration 1:   log pseudolikelihood = -62197.721

Iteration 2:   log pseudolikelihood = -35894.062

Iteration 3:   log pseudolikelihood = -15559.837

Iteration 4:   log pseudolikelihood = -14730.331

Iteration 5:   log pseudolikelihood =  -14720.35

Iteration 6:   log pseudolikelihood = -14720.347

Iteration 7:   log pseudolikelihood = -14720.347


Poisson regression                                Number of obs     =
96

                                                  Wald chi2(10)     =
1141.62

                                                  Prob > chi2       =
0.0000

Log pseudolikelihood = -14720.347                 Pseudo R2         =
0.9034


----------------------------------------------------------------------
--------

             |               Robust

property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

-------------+--------------------------------------------------------
--------

   lnpop2017 |   1.292286   .1386717     9.32   0.000     1.020494
1.564077

   popgrowth |   .1665319   14.54772     0.01   0.991    -28.34648
28.67955

  shareyoung |  -22.99325   12.54162    -1.83   0.067    -47.57437
1.58786

    shareold |   -4.09294   2.953005    -1.39   0.166    -9.880723
1.694842

unemployment |  -.0061053   .0522392    -0.12   0.907    -.1084923
.0962817
```

```
sharepoverty |    4.34683   3.668067     1.19   0.236    -2.842448
11.53611

  highschool |  -.0179484    .032222    -0.56   0.578    -.0811024
.0452056

 somecollege |   .0322355   .0439486     0.73   0.463    -.0539021
.1183731

   associate |    .026788   .0539285     0.50   0.619      -.07891
.132486

  bachelorup |   -.030935   .0246739    -1.25   0.210    -.0792948
.0174249

       _cons |  -4.013571   2.798726    -1.43   0.152    -9.498973
1.47183
----------------------------------------------------------------------
--------


. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty highschool somecollege associate bachelorup,
robust



 Leave-One-Out Cross-Validation Results

-----------------------------------------

         Method          |    Value

-------------------------+---------------

Root Mean Squared Errors |   2644.2267

Mean Absolute Errors     |   715.13287

Pseudo-R2                |   .43527218

-----------------------------------------


. estimate store modelr6


. test somecollege highschool associate
```

```
 ( 1)  [property2017]somecollege = 0
 ( 2)  [property2017]highschool = 0
 ( 3)  [property2017]associate = 0

           chi2(  3) =    0.84
         Prob > chi2 =    0.8404


.
. esttab modelr* using "modelr.ttf" , se(3) replace
(output written to modelr.ttf)


.
. *3.2.5 Estimate the best model
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust


Iteration 0:   log pseudolikelihood = -143334.31
Iteration 1:   log pseudolikelihood =  -78086.38
Iteration 2:   log pseudolikelihood = -20284.874
Iteration 3:   log pseudolikelihood = -15549.804
Iteration 4:   log pseudolikelihood = -15079.904
Iteration 5:   log pseudolikelihood = -15076.717
Iteration 6:   log pseudolikelihood = -15076.717


Poisson regression                               Number of obs     =
96
                                                 Wald chi2(8)      =
768.49
                                                 Prob > chi2       =
0.0000
Log pseudolikelihood = -15076.717                Pseudo R2         =
0.9011
```

```
---------------------------------------------------------------------
--------
             |               Robust
property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+-------------------------------------------------------
--------
   lnpop2017 |   1.318093   .1512191     8.72   0.000     1.021709
1.614477
   popgrowth |   3.042099   14.39551     0.21   0.833    -25.17258
31.25678
  shareyoung |  -21.69349   12.29053    -1.77   0.078    -45.78247
2.395501
    shareold |  -4.202415   3.007807    -1.40   0.162    -10.09761
1.692777
unemployment |   .0147603   .0356234     0.41   0.679    -.0550603
.0845809
sharepoverty |   3.259106   3.537914     0.92   0.357    -3.675078
10.19329
  lessthanhs |  -.0096915   .0197519    -0.49   0.624    -.0484045
.0290215
  bachelorup |  -.0328026   .0256176    -1.28   0.200    -.0830121
.017407
       _cons |  -3.954115   2.013414    -1.96   0.050    -7.900335   -
.0078954
---------------------------------------------------------------------
--------

. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoverty lessthanhs bachelorup, robust

 Leave-One-Out Cross-Validation Results
-----------------------------------------
```

```
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2456.7627
Mean Absolute Errors     |   660.07151
Pseudo-R2                |   .47521604
-----------------------------------------

. estimates store modelz1

.
. poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor, robust

Iteration 0:   log pseudolikelihood = -55584.832
Iteration 1:   log pseudolikelihood = -19654.247
Iteration 2:   log pseudolikelihood = -16015.939
Iteration 3:   log pseudolikelihood = -16007.184
Iteration 4:   log pseudolikelihood = -16007.184

Poisson regression                                Number of obs     =
96
                                                  Wald chi2(7)      =
342.79
                                                  Prob > chi2       =
0.0000
Log pseudolikelihood = -16007.184                 Pseudo R2         =
0.8950

----------------------------------------------------------------------
---------
              |               Robust
 property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
```

```
--------------+----------------------------------------------------------------
---------
    lnpop2017 |   1.033365   .1289532     8.01   0.000     .7806215
1.286109
    popgrowth |   6.721811   12.66734     0.53   0.596    -18.10572
31.54935
   shareyoung |   -18.3381    9.09925    -2.02   0.044     -36.1723
-.5038996
     shareold |  -7.854712    4.49232    -1.75   0.080     -16.6595
.9500728
 unemployment |   .0012247   .0326068     0.04   0.970    -.0626834
.0651327
    sharepoor |   8.274812   3.113321     2.66   0.008     2.172815
14.37681
shareverypoor |   .5404179   5.658764     0.10   0.924    -10.55056
11.63139
        _cons |  -1.811402   2.946822    -0.61   0.539    -7.587068
3.964264
-------------------------------------------------------------------------------
---------


. loocv poisson property2017 lnpop2017 popgrowth shareyoung shareold
unemployment sharepoor shareverypoor, robust



 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2302.1008
Mean Absolute Errors     |   655.54489
Pseudo-R2                |   .51036593
-----------------------------------------
```

```
. estimate store modelz2


.

. poisson property2017 lnpop2017 shareyoung shareold sharepoor
shareverypoor, robust


Iteration 0:   log pseudolikelihood = -53743.698

Iteration 1:   log pseudolikelihood = -19403.733

Iteration 2:   log pseudolikelihood = -16060.876

Iteration 3:   log pseudolikelihood = -16056.429

Iteration 4:   log pseudolikelihood = -16056.429


Poisson regression                                Number of obs     =
96

                                                  Wald chi2(5)      =
245.20

                                                  Prob > chi2       =
0.0000

Log pseudolikelihood = -16056.429                 Pseudo R2         =
0.8946


----------------------------------------------------------------------
---------

              |               Robust

 property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

--------------+-------------------------------------------------------
---------

    lnpop2017 |   1.032816    .128303     8.05   0.000     .7813473
1.284286

   shareyoung |  -18.61604   9.165002    -2.03   0.042    -36.57912
-.6529691

     shareold |  -8.464714   4.542226    -1.86   0.062    -17.36731
.4378849
```

```
   sharepoor |   7.834224   2.876769     2.72   0.006     2.195859
13.47259

shareverypoor |  -.3185715   4.725321    -0.07   0.946    -9.580031
8.942888

       _cons |  -1.495766   2.807232    -0.53   0.594    -6.997841
4.006308
----------------------------------------------------------------------
---------


. loocv poisson property2017 lnpop2017 shareyoung shareold sharepoor
shareverypoor, robust



 Leave-One-Out Cross-Validation Results
-----------------------------------------
        Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2241.3931
Mean Absolute Errors     |   630.28129
Pseudo-R2                |   .53406812
-----------------------------------------


. estimate store modelz3


.
. poisson property2017 lnpop2017 shareyoung shareold sharepoor, robust


Iteration 0:   log pseudolikelihood = -45379.605
Iteration 1:   log pseudolikelihood = -18085.355
Iteration 2:   log pseudolikelihood = -16058.999
Iteration 3:   log pseudolikelihood = -16058.109
Iteration 4:   log pseudolikelihood = -16058.109
```

```
Poisson regression                              Number of obs     =
96

                                                Wald chi2(4)      =
224.03

                                                Prob > chi2       =
0.0000

Log pseudolikelihood = -16058.109               Pseudo R2         =
0.8946

----------------------------------------------------------------------
--------

             |               Robust
property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

-------------+--------------------------------------------------------
--------

   lnpop2017 |   1.032805   .1288242     8.02   0.000     .7803143
1.285296

  shareyoung |  -18.71198   9.628945    -1.94   0.052    -37.58436
.1604101

    shareold |  -8.499526   4.889626    -1.74   0.082    -18.08302
1.083964

   sharepoor |   7.658722    2.89849     2.64   0.008     1.977787
13.33966

       _cons |  -1.484192     2.9412    -0.50   0.614    -7.248837
4.280454

----------------------------------------------------------------------
--------


. loocv poisson property2017 lnpop2017 shareyoung shareold sharepoor,
robust


 Leave-One-Out Cross-Validation Results
```

```
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2083.312
Mean Absolute Errors     |   604.26688
Pseudo-R2                |   .57891546
-----------------------------------------

. estimate store modelz4

.
. poisson property2017 lnpop2017 share14to17 share18to24 shareold
sharepoverty, robust

Iteration 0:   log pseudolikelihood = -70742.098
Iteration 1:   log pseudolikelihood = -29405.873
Iteration 2:   log pseudolikelihood = -16482.542
Iteration 3:   log pseudolikelihood =  -16427.46
Iteration 4:   log pseudolikelihood = -16427.438
Iteration 5:   log pseudolikelihood = -16427.438

Poisson regression                                Number of obs     =
96
                                                  Wald chi2(5)      =
311.30
                                                  Prob > chi2       =
0.0000
Log pseudolikelihood = -16427.438                 Pseudo R2         =
0.8922

----------------------------------------------------------------------
--------
```

```
             |               Robust
property2017 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
-------------+----------------------------------------------------------------
   lnpop2017 |   1.037193   .0893157    11.61   0.000     .8621376
1.212249
 share14to17 |   -16.4947   18.81393    -0.88   0.381    -53.36933
20.37992
 share18to24 |  -20.63601   11.06937    -1.86   0.062    -42.33158
1.059558
    shareold |  -8.411884   4.209881    -2.00   0.046     -16.6631   -
.1606702
sharepoverty |   4.386931   1.842736     2.38   0.017     .7752346
7.998627
       _cons |  -1.523153   2.376442    -0.64   0.522    -6.180893
3.134587
------------------------------------------------------------------------------

. loocv poisson property2017 lnpop2017 share14to17 share18to24
shareold sharepoverty, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   2489.4066
Mean Absolute Errors     |   674.73108
Pseudo-R2                |   .44270051
-----------------------------------------


. estimate store modelz5
```

```
. esttab modelz* using "modelz.rtf" , se(3) replace
file modelz.rtf is read-only; cannot be modified or erased
r(608);

end of do-file

r(608);

. do "C:\Users\luizg\AppData\Local\Temp\STD5764_000000.tmp"

. esttab modelz* using "modelz.rtf" , se(3) replace
(output written to modelz.rtf)

.
. *3.3 Prediction using past crime and other correlates
.
. poisson property2018 propertygrowth lnproperty2017, robust

Iteration 0:   log pseudolikelihood = -5557.0807
Iteration 1:   log pseudolikelihood = -2421.9652
Iteration 2:   log pseudolikelihood = -2398.3012
Iteration 3:   log pseudolikelihood =  -2398.292
Iteration 4:   log pseudolikelihood =  -2398.292

Poisson regression                               Number of obs     =
96
                                                 Wald chi2(2)      =
11022.00
                                                 Prob > chi2       =
0.0000
```

```
Log pseudolikelihood =  -2398.292              Pseudo R2         =
0.9823

----------------------------------------------------------------------
----------
               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+------------------------------------------------------
----------
propertygrowth |   -.504605   .4927611    -1.02   0.306    -1.470399
.461189
lnproperty2017 |    .997609   .0103021    96.84   0.000     .9774172
1.017801
         _cons |  -.1176997   .0754387    -1.56   0.119    -.2655569
.0301574
----------------------------------------------------------------------
----------

. loocv poisson property2018 propertygrowth lnproperty2017, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   468.80571
Mean Absolute Errors     |   134.02932
Pseudo-R2                |   .97684202
-----------------------------------------

. estimates store modelt1
```

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
-------
       Model |          N   ll(null)  ll(model)      df         AIC
BIC
-------------+---------------------------------------------------------------
-------
     modelt1 |         95  -134674.3  -2395.522       3   4797.043
4804.705
-----------------------------------------------------------------------------
-------
Note: BIC uses N = number of observations. See [R] BIC note.

.

. poisson property2018 propertygrowth lnproperty2017 sharepoverty,
robust

Iteration 0:   log pseudolikelihood = -10494.028
Iteration 1:   log pseudolikelihood = -2465.4931
Iteration 2:   log pseudolikelihood = -2398.2678
Iteration 3:   log pseudolikelihood = -2398.2328
Iteration 4:   log pseudolikelihood = -2398.2328

Poisson regression                                Number of obs     =
96
                                                  Wald chi2(3)      =
11148.72
                                                  Prob > chi2       =
0.0000
Log pseudolikelihood = -2398.2328                 Pseudo R2         =
0.9823
```

```
--------------------------------------------------------------------------------
               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+----------------------------------------------------------------
propertygrowth |  -.4928042   .4722747    -1.04   0.297    -1.418446
.4328372
lnproperty2017 |   .9978134   .0104307    95.66   0.000     .9773696
1.018257
  sharepoverty |   .0233358   .3149468     0.07   0.941    -.5939487
.6406202
         _cons |  -.1229764   .1026292    -1.20   0.231    -.3241259
.0781731
--------------------------------------------------------------------------------

. loocv poisson property2018 propertygrowth lnproperty2017
sharepoverty, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   475.7622
Mean Absolute Errors     |   137.38873
Pseudo-R2                |   .97635575
-----------------------------------------

. estimates store modelt2
```

```
. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
       Model |           N   ll(null)  ll(model)      df         AIC         BIC
-------------+---------------------------------------------------------------
     modelt2 |          95  -134674.3  -2395.463       4    4798.926    4809.141
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

.
. poisson property2018 propertygrowth lnproperty2017 sharepoor shareverypoor, robust

Iteration 0:   log pseudolikelihood = -11477.822
Iteration 1:   log pseudolikelihood = -2396.1447
Iteration 2:   log pseudolikelihood = -2308.1802
Iteration 3:   log pseudolikelihood = -2308.1299
Iteration 4:   log pseudolikelihood = -2308.1299

Poisson regression                                Number of obs     =         96
                                                  Wald chi2(4)      =   12103.13
                                                  Prob > chi2       =     0.0000
```

```
Log pseudolikelihood = -2308.1299               Pseudo R2         =
0.9830

----------------------------------------------------------------------
----------

               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]

---------------+------------------------------------------------------
----------

propertygrowth |  -.3604423   .4256824    -0.85   0.397    -1.194764
.4738798

lnproperty2017 |   .9952104   .0098045   101.51   0.000      .975994
1.014427

     sharepoor |     2.1292     1.4918     1.43   0.154    -.7946736
5.053074

 shareverypoor |  -2.338646   1.742944    -1.34   0.180    -5.754753
1.077461

         _cons |  -.1199882   .0911229    -1.32   0.188    -.2985857
.0586094

----------------------------------------------------------------------
----------

. loocv poisson property2018 propertygrowth lnproperty2017 sharepoor
shareverypoor, robust


 Leave-One-Out Cross-Validation Results

-----------------------------------------

         Method          |    Value

-------------------------+---------------

Root Mean Squared Errors |   449.35266

Mean Absolute Errors     |   129.51924

Pseudo-R2                |   .97893602
```

```
-----------------------------------------

. estimates store modelt3

. estat ic

Akaike's information criterion and Bayesian information criterion

-----------------------------------------------------------------------------
        Model |          N   ll(null)  ll(model)      df         AIC         BIC
-------------+---------------------------------------------------------------
      modelt3 |         95  -134674.3  -2305.249       5    4620.497    4633.267
-----------------------------------------------------------------------------
Note: BIC uses N = number of observations. See [R] BIC note.

.
. poisson property2018 propertygrowth lnproperty2017 sharepoor shareverypoor shareyoung, robust

Iteration 0:   log pseudolikelihood = -18254.648
Iteration 1:   log pseudolikelihood =  -2701.798
Iteration 2:   log pseudolikelihood = -2290.7979
Iteration 3:   log pseudolikelihood = -2288.7326
Iteration 4:   log pseudolikelihood = -2288.7326

Poisson regression                              Number of obs     =         96
```

```
                                                Wald chi2(5)      =
9116.47
                                                Prob > chi2       =
0.0000
Log pseudolikelihood = -2288.7326               Pseudo R2         =
0.9831

----------------------------------------------------------------------
----------
               |               Robust
  property2018 |      Coef.   Std. Err.      z    P>|z|     [95% Conf.
Interval]
---------------+------------------------------------------------------
----------
propertygrowth |  -.5020992   .3833684    -1.31   0.190    -1.253488
.2492891
lnproperty2017 |   .9954931   .0107959    92.21   0.000     .9743336
1.016653
     sharepoor |   2.302194   1.446691     1.59   0.112    -.5332676
5.137657
 shareverypoor |  -2.630776   1.675436    -1.57   0.116    -5.914571
.6530185
    shareyoung |   1.284683   .7946298     1.62   0.106     -.272763
2.842128
         _cons |  -.3091024   .1278282    -2.42   0.016    -.5596412
-.0585637
----------------------------------------------------------------------
----------

. loocv poisson property2018 propertygrowth lnproperty2017 sharepoor
shareverypoor shareyoung, robust


 Leave-One-Out Cross-Validation Results
-----------------------------------------
```

```
         Method          |    Value
-------------------------+---------------
Root Mean Squared Errors |   505.79555
Mean Absolute Errors     |   136.32802
Pseudo-R2                |   .97625669
-----------------------------------------


. estimates store modelt4


. estat ic


Akaike's information criterion and Bayesian information criterion


----------------------------------------------------------------------
-------
       Model |          N   ll(null)  ll(model)      df         AIC
BIC
-------------+--------------------------------------------------------
-------
     modelt4 |         95  -134674.3  -2285.781       6   4583.563
4598.886
----------------------------------------------------------------------
-------
Note: BIC uses N = number of observations. See [R] BIC note.


. esttab modelt* using "modelt.rtf" , se(3) replace
(output written to modelt.rtf)


.
. *3.4 Predictions for 2018
. gen lnproperty2018 = ln(property2018)
```

```
. gen CrimeRate = (lnproperty2018-lnproperty2012)/6


.

. gen property2019=exp(-
0.364423*propertygrowth+0.9952104*lnproperty2018+2.1292*sharepoor-
2.338646*shareverypoor-0.1199882)


.

. list county property2019 (3), compress

3 invalid name

r(198);


end of do-file


r(198);


. do "C:\Users\luizg\AppData\Local\Temp\STD5764_000000.tmp"


. list county property2019, compress


     +-----------------------+

     |     county   pro~2019 |

     |-----------------------|

  1. |    Baldwin   493.0656 |

  2. |      Banks   344.8884 |

  3. |   Ben Hill   159.2014 |

  4. |    Berrien   197.3487 |

  5. |   Bleckley   92.02151 |

     |-----------------------|

  6. |     Brooks   125.4411 |

  7. |      Bryan   291.8895 |
```

8.	Bulloch	439.9952
9.	Butts	289.4147
10.	Charlton	103.44
11.	Chattooga	81.77737
12.	Cherokee	1174.619
13.	Clay	10.23722
14.	Clayton	7472.627
15.	Clinch	40.65041
16.	Cobb	8181.151
17.	Coffee	435.7638
18.	Columbia	1589.328
19.	Cook	119.8019
20.	Coweta	901.6461
21.	Crawford	181.4027
22.	Crisp	399.5795
23.	Dawson	394.2889
24.	Decatur	191.386
25.	DeKalb	19157.35
26.	Dodge	173.2219
27.	Dooly	81.68842
28.	Dougherty	368.4775
29.	Early	95.41566
30.	Echols	46.39896
31.	Effingham	296.6019
32.	Elbert	258.4436

```
33. |    Emanuel   209.9395 |
34. |      Evans   32.68354 |
35. |     Fannin   187.0679 |
    |-----------------------|
36. |    Fayette   456.6325 |
37. |     Fulton   1285.335 |
38. |     Gilmer   386.3245 |
39. |      Glynn   1506.552 |
40. |      Grady   166.4527 |
    |-----------------------|
41. |     Greene   104.7981 |
42. |   Gwinnett   12713.54 |
43. |  Habersham    316.316 |
44. |       Hall   1445.785 |
45. |    Hancock   91.26628 |
    |-----------------------|
46. |      Henry   3129.087 |
47. |    Houston   820.4178 |
48. |      Irwin   93.40747 |
49. |     Jasper   152.3876 |
50. | Jeff Davis   93.18154 |
    |-----------------------|
51. |  Jefferson   75.20541 |
52. |      Jones   225.8053 |
53. |      Lamar   99.58595 |
54. |     Lanier   129.8812 |
55. |    Laurens   452.0814 |
    |-----------------------|
56. |        Lee   466.0869 |
57. |    Liberty   203.4765 |
```

58.	Long	143.9327
59.	Madison	365.3856
60.	Mitchell	229.7884
61.	Monroe	274.5731
62.	Newton	1382.42
63.	Oconee	400.3466
64.	Oglethorpe	172.1628
65.	Paulding	1759.773
66.	Peach	201.4963
67.	Pickens	371.4142
68.	Pierce	237.222
69.	Pike	178.0229
70.	Polk	506.5225
71.	Pulaski	197.3949
72.	Rabun	190.5708
73.	Rockdale	1120.019
74.	Schley	9.490532
75.	Seminole	53.3412
76.	Spalding	748.7526
77.	Stephens	242.0559
78.	Sumter	172.5431
79.	Talbot	40.73468
80.	Taylor	41.41781
81.	Thomas	427.7515
82.	Towns	123.9782

```
 83. |   Treutlen   40.05066 |
 84. |      Troup   569.7014 |
 85. |     Twiggs   92.75941 |
     |-----------------------|
 86. |      Union   197.3873 |
 87. |      Upson   265.1551 |
 88. |     Walton   519.6805 |
 89. |       Ware   529.1973 |
 90. |     Warren   27.24125 |
     |-----------------------|
 91. |      Wayne   628.5408 |
 92. |    Webster   1.182196 |
 93. |      White   122.7599 |
 94. |  Whitfield   1368.869 |
 95. |     Wilkes    18.5601 |
     |-----------------------|
 96. |  Wilkinson   34.49257 |
     +-----------------------+

.
. STOP
command STOP is unrecognized
r(199);

end of do-file

r(199);

.
```