INTRODUCTION AND SELECTION OF VARIABLES

Historically, crime rate has been determined by deterrence measures, economic conditions, demographics, and cultural or family background. Most data indicates that deterrence measures have the most significant negative effect on crime, especially the probability that an individual will be arrested, convicted, and imprisoned (Baltagi, 2006). The number of police officials also serves as an important deterrent (Corman & Mocan, 2000). Other factors such as high inequality, poverty, and proportion of young males positively affect the crime rate (Kelly, 2000). The population density might also impact the crime rate (Levitt, 2004).

INITIAL MODEL AND CRITIQUE

For my initial model, I chose deterrence variables *prbarr* (probability of arrest), *prbconv* (probability of conviction if arrested), *prbpris* (probability of prison sentence if convicted), *avgsen* (average prison sentence in days), and *polpc* (police per capita). While looking at the scatterplot of *crimerate* and *polpc*, I noticed an outlier in *polpc* (Fig. 1) and used a dummy variable (*polpcdum*) to account for it. I included *density* as it seemed important according to theory and had a high correlation (0.719) with the dependent variable *crimerate*. To represent the effect of race and age, I included *pctmin80* (percentage of ethnic minority in 1980) and *pctymle* (percentage of the population that is young male). Lastly, labour market conditions were reflected by *wmfg* (the weekly wages in manufacturing sector). I chose to not use logs initially as most variables lie between zero and one.

The Model

$$\begin{aligned}\widehat{crimerate} = {} & \underset{(0.0089158)}{\mathbf{0.0030489}} - \underset{(0.0089158)}{\mathbf{0.020587}\,prbarr} - \underset{(0.005265)}{\mathbf{0.0124127}prbconv} \\ & + \underset{(0.0150414)}{\mathbf{0.0263064}prbpris} - \underset{(0.000474)}{\mathbf{0.0007114}avgsen} - \underset{(0.5194857)}{\mathbf{0.8370832}polpc} \\ & + \underset{(0.0206611)}{\mathbf{0.0658}polpcdum} + \underset{(0.0008368)}{\mathbf{0.0067489}density} + \underset{(0.0000618)}{\mathbf{0.0001755}pctmin80} \\ & + \underset{(0.000014)}{\mathbf{0.0000292}wmfg} + \underset{(0.0454569)}{\mathbf{0.1298684}pctymle}\end{aligned}$$

N=89, R^2=0.7290, adj. R^2=0.6943

Analysis and Critique

As suggested by theory, the coefficients for deterrence variables were negative except for *prbpris*. The coefficients of *density*, *pctmin80*, and *pctymle* were also positive as predicted. The model indicated that an increase in *wmfg* led to an increase in *crimerate*.

Most variables were significant at 5% except for *prbpris*, *avgsen,* and *polpc*. Conducting an F-test for *polpc* and *polpcdum* gave a p-value of 0.0008, indicating that they were jointly very significant. The magnitude of the coefficients of *avgsen* and *pctmin80* were much smaller than other variables. The magnitude of the coefficient of *wmfg* was also relatively small but that may be due to its higher values (ranging from 145.9 to 576.33). The model seemed to be a good fit with an R^2 of 0.7290.

I have assumed that the data provided satisfies Multiple Linear Regression (MLR) assumption 2 of random sampling (see appendix). There was no perfect collinearity suggested by either theory or data, although the large number of variables may have caused some multicollinearity. The Variance Inflation Factors (VIF) (value 1.93) indicated that there was little multicollinearity in this model. Ramsey's RESET did not indicate functional form misspecification (p-value 0.7269). Given the data, none of the regressors appeared to be highly correlated with variables that might have ended up in the error term.

On the other hand, White's test suggested that the model possessed unrestricted heteroskedasticity (the null hypothesis of homoskedasticity is rejected even at 1% significance). The error terms were not normally distributed and were skewed towards the right.

The model appeared to satisfy MLR assumptions 1-4 but not MLR 5-6. Thus, the estimators were likely not biased but may have been inconsistent. The sample size (89) is large enough to assume asymptotic normality and accept results from t- and F-tests. The R^2 and adj. R^2 were reasonably high (0.7290 and 0.6943 respectively) but this may have been because of the large number of variables.

IMPROVEMENTS TO FIRST MODEL

I replaced *wmfg* with its logarithm *log_wmfg*. Since most wage variables were correlated with each other, I used only the most significant one. In the initial model, *avgsen* was insignificant at 5% level and its coefficient was comparatively smaller in magnitude. I removed *avgsen* and *pctmin80* after trying different functional forms as they were insignificant at 5%. I added *west* (=1 if the individual lives in Western NC) and replaced *density* with *urban* (=1 if in SMSA). These changes considerably reduced heteroskedasticity in my model as per White's test (p-value 0.1254).

I tested the squared terms of all deterrence variables and found the square of *prbconv* to be highly significant. *Prbpris* and its squared term were both individually and jointly insignificant at usual levels. The square of *prbarr* was also insignificant. Adding the square of *prbconv* made *prbpris* and *polpc* much less significant, so I removed them from the regression. The model without either of these variables performed better on Ramsey's RESET, although the p-value for the same went down considerably compared to the initial model.

SECOND MODEL AND CRITIQUE

The Model

$$\widehat{crimerate} = \underset{(0.0280711)}{-0.0419896} \underset{(0.009151)}{-0.0371891 prbarr} + \underset{(0.0214622)}{0.0506352 prbconv} - \underset{(0.0150414)}{0.0263064 prbconv2} + \underset{(0.0045062)}{0.0243091 urban} + \underset{(0.0047763)}{0.0122224 logw\_mfg} + \underset{(0.0488497)}{0.0934561 pctymle} - \underset{(0.0026271)}{0.0111573 west}$$

N=89, R^2= 0.6513, adj. R^2= 0.6211

Analysis and Critique

All variables were significant at 5% significance except for *pctymle* (p-value of 0.059 on the t-test). I still included it as removing it from the regression drastically decreased both the p-value for Ramsey's RESET (from 0.1788 to 0.1065) and White's test (from 0.1913 to 0.0279).

The residuals for this model were slightly more normal, although the skewness and kurtosis test (sktest) still rejected normality at usual significance levels (p-value 0.002). Despite this, the model appeared to have much less heteroskedasticity than before. White's test could not reject the null hypothesis of homoskedasticity even at 10% significance. The second model seemed to satisfy MLR assumptions 1-5, but there may have been a better functional form. Further, although there was no perfect collinearity (MLR 3), there was some multicollinearity due to the presence of both *prbconv* and its squared term (VIF 5.24). This was still at an acceptable level.

To improve the functional form of my model, I transformed *crimerate* into its logarithmic form *log_CR*. I also tried using the lagged *crimerate* variable. Although theoretically and statistically

important, it was highly correlated with most of the other regressors (in particular, correlation between *crmrtelag* and *prbarr* is -0.4797) and did worse on both RESET and White's test. It would also have led to overcontrolling.

FINAL MODEL AND CRITIQUE

The Model

$$\widehat{log\_CR} = \underset{(1.009032)}{-6.357204} \underset{(0.3289396)}{-1.649196\,prbarr} + \underset{(0.7714725)}{2.766648prbconv}$$
$$[1.316318] \quad [1.316318] \quad [0.9559553]$$
$$\underset{(0.7752083)}{-3.47633prbconv2} + \underset{(0.1619771)}{0.4778541urban} + \underset{(0.171688)}{0.4618877logw\_mfg}$$
$$[0.8441244] \quad [0.0994015] \quad [0.2262556]$$
$$\underset{(1.755931)}{+2.595302pctymle} - \underset{(0.0944344)}{0.4519147west}$$
$$[1.161621] \quad [0.0932804]$$

N=89, R^2= 0.6326

Analysis and Critique

When using heteroskedasticity robust standard errors, all variables were significant at 5% level. With respect to regular standard errors, *pctymle* was insignificant at 10%. I decided to keep it in the model due to its significance while using robust errors as well as a better adj. R^2, lower AIC, and more homoskedasticity than when it was not present.

Transforming *crimerate* changed the p-value for Ramsey's RESET drastically (from 0.1788 to 0.9160, which is extremely high). Moreover, the p-value for White's test increased to 0.5457. Thus, it was a better version of the previous model. Nevertheless, the sktest still rejected normality with a p-value of 0.0025.

This was my preferred specification as it included all variables that were significant but not highly correlated with other significant variables. It overwhelmingly rejected heteroskedasticity and functional form misspecification and satisfied MLR 1-5. It also had a relatively high R^2 (0.6326), indicating that the regressors explain around 63% of the variation in *crimerate*. There was some multicollinearity (VIF 5.24) but reducing it would have increased heteroskedasticity. The error terms were still not normal but could be considered asymptotically normal due to the large sample size.

The final model might have suffered from omitted variable bias as there was no data for theoretically important variables such as family background, unemployment, poverty, and drug use. There might also have been a simultaneity bias as crime rate and probability of arrest both affect each other simultaneously. However, this cannot be tested without additional data.

Economic Interpretation of the Model

The model indicates that the crime rate in North Carolina in 1985 is a function of deterrence, demographic, and labour market variables. A 0.01 (1%) increase in the probability of arrest decreases the elasticity of crime rate by 1.65. An increase in the probability of conviction initially has diminishing marginal returns to the elasticity of the rate of crime, after which log of crime rate

decreases as probability of conviction increases. This supports the hypothesis that deterrence variables have a negative relationship with the rate of crime and have the most impact.

The demographic variables indicate that percentage of crime rate increases by 47.78% in SMSAs and decreases by 45.19% if in Western North Carolina. Moreover, percentage of crime rate is 2.59% higher for every 0.01 unit (1%) increase in the proportion of young males in the population. This is in accordance with the fact that young males are more predisposed to commit crime compared to their adult counterparts (Kelly, 2000).

Finally, a 1% increase in wages in the manufacturing sector increase the percentage of crime rate by 0.46%. This might be due to high inequality—placing poor individuals next to high income individuals increases returns to crime which increases crime rate (Kelly, 2000). This can be verified by including unemployment or other measures of poverty in the regression.

(Word count including appendix: 1495)

BIBLIOGRAPHY

Baltagi, B. H., 2006. Estimating an Economic Model of Crime Using Panel Data From North Carolina. *Journal of Applied Econometrics,* Volume 21, pp. 543-547.

Becker, G. S., 1968. Crime and Punishment: An Economic Approach. *The Journal of Political Economy,* 76(2), pp. 169-217.

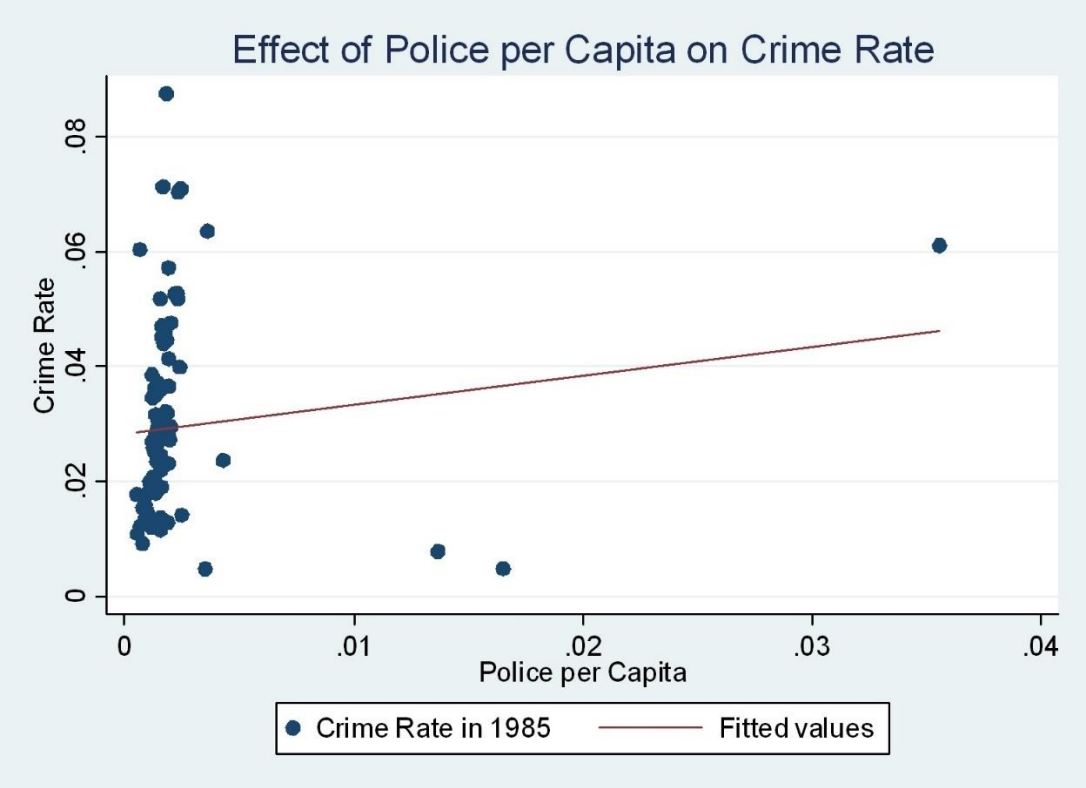
Corman, H. & Mocan, H. N., 2000. A Time-Series Analysis of Crime, Deterrence, and Drug Abuse in New York City. *The American Economic Review,* 90(3), pp. 584-604.

Kelly, M., 2000. Inequality and Crime. *The Review of Economics and Statistics,* 82(4), pp. 530-539.

Levitt, S. D., 2004. Understanding Why Crime Fell in the 1990s: Four Factors that Explain the Decline and Six that Do Not. *Journal of Economic Perspectives,* 18(1), pp. 163-190.

APPENDIX

Figure 1: Scatterplot between *crimerate* and *polpc*



Stata Output for Initial Model

Source	SS	df	MS	Number of obs	=	89
				F(10, 78)	=	20.98
Model	.018431626	10	.001843163	Prob > F	=	0.0000
Residual	.006851383	78	.000087838	R-squared	=	0.7290
				Adj R-squared	=	0.6943
Total	.025283009	88	.000287307	Root MSE	=	.00937

crimerate	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
prbarr	-.020587	.0089158	-2.31	0.024	-.038337	-.002837
prbconv	-.0124127	.005265	-2.36	0.021	-.0228945	-.0019308
prbpris	.0263064	.0150414	1.75	0.084	-.0036388	.0562515
avgsen	-.0007114	.000474	-1.50	0.137	-.001655	.0002322
polpc	-.8370832	.5194857	-1.61	0.111	-1.8713	.1971334
polpcdum	.0658	.0206611	3.18	0.002	.0246669	.1069331
density	.0067489	.0008368	8.07	0.000	.005083	.0084148
pctmin80	.0001755	.0000618	2.84	0.006	.0000525	.0002986
wmfg	.0000292	.000014	2.08	0.041	1.28e-06	.0000572
pctymle	.1298684	.0454569	2.86	0.005	.0393707	.2203661
_cons	.0030489	.0112478	0.27	0.787	-.0193437	.0254416

Multiple Linear Regression Assumptions

1. Linear in parameters and correctly specified
2. Random sample
3. No perfect collinearity
4. Zero conditional mean
5. Homoskedasticity
6. Normality of error terms

Stata Output for Second Model

Source	SS	df	MS	Number of obs	=	89
				F(7, 81)	=	21.61
Model	.016466458	7	.002352351	Prob > F	=	0.0000
Residual	.008816552	81	.000108846	R-squared	=	0.6513
				Adj R-squared	=	0.6211
Total	.025283009	88	.000287307	Root MSE	=	.01043

crimerate	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
prbarr	-.0371891	.009151	-4.06	0.000	-.0553968	-.0189814
prbconv	.0506352	.0214622	2.36	0.021	.0079321	.0933383
prbconv2	-.0685145	.0215662	-3.18	0.002	-.1114244	-.0256046
urban	.0243091	.0045062	5.39	0.000	.0153432	.0332749
log_wmfg	.0122224	.0047763	2.56	0.012	.002719	.0217258
pctymle	.0934561	.0488497	1.91	0.059	-.0037395	.1906517
west	-.0111573	.0026271	-4.25	0.000	-.0163845	-.0059301
_cons	-.0419896	.0280711	-1.50	0.139	-.0978423	.0138631

Stata Output for Final Model

Linear regression

Number of obs	=	89
F(7, 81)	=	49.52
Prob > F	=	0.0000
R-squared	=	0.6326
Root MSE	=	.37502

log_CR	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
prbarr	-1.649196	.3142446	-5.25	0.000	-2.274444	-1.023948
prbconv	2.766648	.9559553	2.89	0.005	.8645972	4.6687
prbconv2	-3.47633	.8441244	-4.12	0.000	-5.155873	-1.796788
urban	.4778541	.0994015	4.81	0.000	.2800763	.675632
log_wmfg	.4618877	.2262556	2.04	0.044	.0117101	.9120653
pctymle	2.595302	1.161621	2.23	0.028	.2840402	4.906565
west	-.4519147	.0932804	-4.84	0.000	-.6375135	-.2663159
_cons	-6.357204	1.316318	-4.83	0.000	-8.976265	-3.738144

Project .do File

```
capture log close

cd "H:\Econometrics\EC3301 project"

use projectcrime.dta, clear

log using "ProjectLog.log", replace

/*labelling variables*/

label data "Crime Rate and Related Variables by County, North Carolina, 1985"

label var county "County"

label var crimerate "Crime Rate in 1985"

label var prbarr "Probability of Arrest"

label var prbconv "Probability of Conviction if Arrested"

label var prbpris "Probability of Prison Sentence if Convicted"

label var avgsen "Average Prison Sentence in Days"

label var polpc "Police per Capita"

label var density "Number of People per Square Mile"

label var taxpc "Tax Revenue per Capita"

label var west "(=1) if Living in Western NC"
```

```
label var central "(=1) if Living in Central NC"

label var urban "(=1) if Living in a Standard Metropolitan Statistical
Area"

label var pctmin80 "Percentage of Ethnic Minority in 1980"

label var wcon "Weekly Wage in Construction in Dollars"

label var wtuc "Weekly Wage in Transport, Utilities, and Communications
in Dollars"

label var wtrd "Weekly Wage in Wholesale or Retail Trade in Dollars"

label var wfir "Weekly Wage in Financial Services, Insurance, and Real
Estate in Dollars"

label var wser "Weekly Wage in the Service Industry in Dollars"

label var wmfg "Weekly Wage in Manufacturing in Dollars"

label var wfed "Weekly Wage of Federal Employees in Dollars"

label var wsta "Weekly Wage of State Employees in Dollars"

label var wloc "Weekly Wage of Local Government Employees in Dollars"

label var pctymle "Young Males as a Percentage of Population"

label var crmrtelag "Crime Rate in North Carolina in 1984"

/*looking at the dataset*/

browse

/*getting a summary for the data*/

summarize

/*Notice that wcon has an extremely high maximum value compared to its
minimum, average, and maximum for wages in other industries. Similarly,
pctymle has a lower mean than expected.*/

scatter crimerate wcon

scatter crimerate pctymle

/*As predicted, there is a large outlier for both of these variables.*/

/*using the correlation matrix to identify key relationships between
variables according to the dataset*/

correlate

/*we can see that urban and density are highly correlated. the wage
variables also seem to be correlated. can reasonably assume that most
wages move together*/
```

```
/*using scatterplots to identify outliers in chosen variables*/

scatter crimerate prbarr || lfit crimerate prbarr

/*seems to fit reasonably well*/

scatter crimerate prbconv || lfit crimerate prbconv

/*seems to fit reasonably well*/

scatter crimerate prbpris || lfit crimerate prbpris

/*seems to fit reasonably well*/

scatter crimerate avgsen || lfit crimerate avgsen

/*seems to fit reasonably well*/

scatter crimerate polpc || lfit crimerate polpc , title("Effect of
Police per Capita on Crime Rate") , xtitle("Police per Capita") ,
ytitle("Crime Rate")

/*there may be a potential outlier for polpc>0.03*/

scatter crimerate density || lfit crimerate density

/*seems to fit reasonably well*/

scatter crimerate pctmin80 || lfit crimerate pctmin80

/*seems to fit reasonably well*/

scatter crimerate wmfg || lfit crimerate wmfg

/*seems to fit reasonably well*/

scatter crimerate pctymle || lfit crimerate pctymle

/*there may be a potential outlier for pctymle>0.25*/

/*trying out initial regression without adjusting for outliers. I will
adjust functional form based on distrubution of errors.*/

generate polpcdum=0

replace polpcdum=1 if polpc>0.03

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 wmfg pctymle

estat vif

ovtest

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 wmfg pctymle

predict res, r
```

```
histogram res, normal

sktest res

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 wmfg pctymle

estat imtest, white

hettest

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 wmfg pctymle

test polpc polpcdum

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 wmfg pctymle, vce(robust)

/*comparing with and without wmfg. adding different wages and
checking*/

generate log_wmfg=log(wmfg)

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 log_wmfg pctymle

estat imtest, white

generate log_wtuc=log(wtuc)

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 log_wtuc pctymle

generate log_wtrd=log(wtrd)

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum density
pctmin80 log_wtrd pctymle

/*wmfg seems to fit the best in its logged form.*/

/*trying white's test with urban in place of density*/

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum urban
pctmin80 log_wmfg pctymle

estat imtest, white

/*does better on the white test. keeping this change.*/

generate log_avgsen=log(avgsen)

regress crimerate prbarr prbconv prbpris avgsen polpc polpcdum urban
pctmin80 log_wmfg pctymle

regress crimerate prbarr prbconv prbpris log_avgsen polpc polpcdum
urban pctmin80 log_wmfg pctymle
```

```
/*log of avgsen is not very significant, and avgsen is also not
significant at 5%.*/

/*avgsen seems to be insignificant at 5%. trying a quadratic approach*/

generate avgsen2=avgsen^2

regress crimerate prbarr prbconv prbpris avgsen avgsen2 polpc polpcdum
urban pctmin80 log_wmfg pctymle

/*both variables are insignificant. removing avgsen from the
regression*/

regress crimerate prbarr prbconv prbpris polpc polpcdum urban pctmin80
log_wmfg pctymle

/*pctmin80 has an extremely small coefficient. trying without*/

regress crimerate prbarr prbconv prbpris polpc polpcdum urban log_wmfg
pctymle

/*trying west and central*/

regress crimerate prbarr prbconv prbpris polpc polpcdum urban log_wmfg
pctymle west central

/*west is very significant, central is not. removing central*/

regress crimerate prbarr prbconv prbpris polpc polpcdum urban log_wmfg
pctymle west

estat imtest, white

/*passes the white test*/

regress crimerate prbarr prbconv prbpris polpc polpcdum urban log_wmfg
pctymle west

ovtest

/*trying squares of deterrence variables*/

generate prbarr2 = prbarr^2

generate prbconv2 = prbconv^2

generate prbpris2 = prbpris^2

regress crimerate prbarr prbconv prbpris polpc polpcdum urban log_wmfg
pctymle west

regress crimerate prbarr prbarr2 prbconv prbconv2 prbpris prbpris2
polpc polpcdum urban log_wmfg pctymle west

test prbarr prbarr2

test prbpris prbpris2
```

```
test prbconv prbconv2

regress crimerate prbarr prbconv prbconv2 prbpris polpc polpcdum urban
log_wmfg pctymle west

test prbconv prbconv2

/*keep only square of prbconv*/

regress crimerate prbarr prbconv prbconv2 polpc polpcdum urban log_wmfg
pctymle west

estat imtest, white

test polpc polpcdum

/*polpc is much less significant. testing without*/

regress crimerate prbarr prbconv prbconv2 polpc polpcdum urban log_wmfg
pctymle west

ovtest

regress crimerate prbarr prbconv prbconv2 urban log_wmfg pctymle west

ovtest

estat imtest, white

estat vif

regress crimerate prbarr prbconv prbconv2 prbpris urban log_wmfg
pctymle west

ovtest

estat imtest, white

estat vif

/*RESET is much better for the second model. although first model is
less heteroskedastic, both pass white's test*/

/*trying taxpc*/

regress crimerate prbarr prbconv prbconv2 urban log_wmfg pctymle west
taxpc

predict res1, r

histogram res1, normal

sktest res1

/*there is still skewness and kurtosis but much better model than
before. taxpc is very insignificant. removing from regression*/
```

```
regress crimerate prbarr prbconv prbconv2 urban log_wmfg pctymle west
predict res2, r
histogram res2, normal
sktest res2
regress crimerate prbarr prbconv prbconv2 urban log_wmfg pctymle west
estat vif
ovtest
estat imtest, white
/*deciding whether to keep pctymle*/
regress crimerate prbarr prbconv prbconv2 urban log_wmfg pctymle west
ovtest
estat imtest, white
regress crimerate prbarr prbconv prbconv2 urban log_wmfg west
ovtest
estat imtest, white
/*misspecification increases by a lot without it. keeping it for now.*/
/*trying log of crimerate*/
generate log_CR = log(crimerate)
regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west
predict res3, r
histogram res3, normal
sktest res3
regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west
estat vif
ovtest
estat imtest, white
/*a better model - better RESET/White's test and normality is the
same.*/
/*trying to add lagged crimerate*/
generate log_LCR=log(crmrtelag)
```

```
regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west
log_LCR

estat vif

ovtest

estat imtest, white

predict res4, r

histogram res4, normal

sktest res4

test prbconv prbconv2

correlate prbarr crmrtelag

/*prbarr is insignificant. this is because prbarr and lagged crimerate
are reasonably correlated. trying without*/

regress log_CR prbconv prbconv2 urban log_wmfg pctymle west log_LCR

ovtest

estat imtest, white

test prbconv prbconv2

estat vif

predict res5, r

histogram res5, normal

sktest res5

regress log_CR prbconv prbconv2 density log_wmfg pctymle west log_LCR

ovtest

estat imtest, white

test prbconv prbconv2

estat vif

/*density is significant but this model has a worse specification and
more heteroskedasticity*/

generate log_dens=log(density)

regress log_CR prbconv prbconv2 log_dens log_wmfg pctymle west log_LCR

ovtest
```

```
estat imtest, white

/*does much worse on heteroskedasticity. not keeping density*/

regress log_CR prbconv prbconv2 log_wmfg pctymle west log_LCR

ovtest

estat imtest, white

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west

ovtest

estat imtest, white

/*removing urban makes it much worse. keeping second specification.*/

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west

ovtest

estat imtest, white

estat ic

regress log_CR prbarr prbconv prbconv2 urban log_wmfg west

ovtest

estat imtest, white

estat ic

/*better R squared, better White's test, similar RESET and lower AIC if
pctymle is included. BIC lower for the second model and it has slightly
better specification.*/

/*trying log of prbconv instead of quadratic*/

generate log_pconv=log(prbconv)

regress log_CR prbarr log_pconv urban log_wmfg pctymle west

estat vif

ovtest

estat imtest, white

/*makes it very heteroskedastic. sticking to quadratics*/

generate log_wcon=log(wcon)

scatter crimerate log_wcon

generate log_wcondum=0
```

```
replace log_wcondum=1 if log(wcon)>7

regress log_CR prbarr prbconv prbconv2 urban log_wmfg log_wcon
log_wcondum pctymle west

estat vif

ovtest

estat imtest, white

/*not very significant; makes heteroskedasticity much worse*/

regress log_CR prbarr prbconv prbconv2 urban log_wcon log_wcondum
pctymle west

estat vif

ovtest

estat imtest, white

estat ic

/*does worse on White's test. much higher AIC/BIC. keeping only wmfg*/

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west

test prbconv prbconv2

estat vif

ovtest

estat imtest, white

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west

predict res6, r

histogram res6, normal

sktest res6

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west

regress log_CR prbarr prbconv prbconv2 urban log_wmfg pctymle west,
vce(robust)

estat vif

ovtest

estat imtest, white

predict res7, r

histogram res7, normal
```

```
sktest res7


log close
```