Group 14:

Comments and Explanation

__Comment 1:__ ***You should write out a formal equation of your regression, as discussed in class many times. Also note that all these regressors are endogenous.***

Predicting crime rates (criv) using various predictors (pris, state, black, unem, polpc, metro).

Formal Regression Equation (this equation is a starting point for analysis, used for theoretical discussions, and not the actual model employed in the research. Model 1 is the actual one) :

$criv = \beta_0 + \beta_1 \times pris + \beta_2 \times state + \beta_3 \times black + \beta_4 \times unem + \beta_5 \times polpc + \beta_6 \times metro + \epsilon$

Where:

- criv is the crime rate.
- pris, state, black, unem, polpc, and metro are the independent variables.
- $\beta_0, \beta_1, \ldots, \beta_6$ are the coefficients to be estimated.
- ϵ represents the error term.

__Comment 2:__ ***You should explain this further, but at least you pointed it out.***

Further Analysis: Addressing Endogeneity in the Relationship Between Crime Rates and Prison Populations

In our initial analysis of the correlation between crime rates and prison populations using panel data, we uncovered significant associations that provide preliminary insights into the dynamics of crime and incarceration. However, to elevate our understanding from correlation to causation—essential for informing effective policy interventions—we must address complex statistical issues such as reverse causality and omitted variable bias.

Reverse Causality:

Reverse causality refers to a scenario where the presumed cause and effect relationship is bidirectional rather than unidirectional. In the context of our study, while it might be hypothesized that increases in prison populations lead to reductions in crime rates (presumably through a deterrent or incapacitation effect), there is also a plausible reverse pathway: rising crime rates could lead to higher prison populations as more individuals are arrested and convicted. This possibility suggests that our observed correlation might not straightforwardly reflect the impact of incarceration on crime, but rather an intertwined relationship where each influences the other.

Omitted Variable Bias:

Omitted variable bias arises when a model fails to include one or more relevant variables that influence both the dependent variable (crime rates) and one or more independent variables (prison population). The absence of these variables from the model can distort the estimated effect of the independent variables on the dependent variable. For instance, factors such as the effectiveness of law enforcement, socioeconomic conditions, and local policies may simultaneously influence both crime rates and the size of the prison population. Ignoring these factors could lead us to erroneously attribute changes in crime rates solely to changes in prison populations.

Comment 3 and 4: ***I do not see a clear description of these variables anywhere. In the data section it says they represent a final decision on litigation?? Idk what that means. If you are going to use these as instruments, you should clearly explain what they are.***

I still don't understand what's going on in this regression. You need to explain why these would be good instruments. Is i a state? An individual? A prison? Having a final decision on litigation affects prison population directly without directly influencing violent crimes?

Explanation of Instrumental Variables for the Regression Model

Context of the Model:

The regression model is developed to estimate factors influencing the growth in the prison population (gpris_it) across different states (denoted by i). This analysis uses panel data that captures variations over time within each state.

Instrumental Variables Explanation:

final1: This binary indicator equals 1 if there was a final decision on litigation in the current year within the state. It is intended to capture the immediate impact of litigation outcomes, such as sentencing or case closures, on the state's prison population.

final2: This variable equal 1 if there was a decision on litigation in the previous two years. It provides a broader measure of the judicial decisions' impact over a slightly extended period, reflecting slower changes in the prison system due to legal proceedings.

Justification for Using final1 and final2 as Instruments:

Relevance: Both variables are directly related to judicial activities that can lead to incarceration. By capturing the rate at which litigation is resolved conclusively, these indicators serve as proxies for the flow of individuals entering the prison system.

Exogeneity: The key assumption here is that while these litigation outcomes affect prison admissions, they are unlikely to be directly influenced by the current crime rates themselves, thus meeting the requirement for an instrumental variable. They are more likely influenced by legal frameworks, policy changes, or efficiency of the court systems, which are independent of the crime occurrences in the near term.

Clarifying the Role of "final1" and "final2":

The inclusion of "final1" and "final2" in the model addresses potential endogeneity between the observed prison population growth and unobserved factors that could simultaneously influence crime rates and prison admissions. By using these as instruments, the model aims to isolate the effect of changes in litigation outcomes from other variables that directly impact crime statistics.

Addressing Potential Confusions:

State-Level Analysis: It's important to clarify that i represents states, not individual prisons or persons. This distinction is critical as it affects the interpretation of how statewide policies and judicial decisions impact prison populations.

Impact on Violent Crimes: While these variables help explain changes in the prison population, they do not directly influence violent crimes, thereby justifying their use as instruments rather than direct predictors in a model focusing on crime rates.

<u>Comment 5</u>: *Why aren't you showing any tables of your coefficient estimates? I discussed in class multiple times how I wanted you to present your estimates.*

Model 1 (gcriv by OLS):

```
Residuals:
     Min       1Q   Median       3Q      Max 
-0.32282 -0.04186  0.00283  0.04109  0.50580 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.005671   0.021558  -0.263 0.792597    
gpris       -0.180897   0.047628  -3.798 0.000159 ***
gpolpc       0.051424   0.055516   0.926 0.354619    
gincpc       0.738368   0.166378   4.438 1.06e-05 ***
cunem        0.411260   0.393670   1.045 0.296536    
cblack      -0.014743   0.033157  -0.445 0.656708    
cmetro       0.538306   0.995660   0.541 0.588922    
cag0_14      0.989306   2.006539   0.493 0.622140    
cag15_17     4.983840   4.740475   1.051 0.293472    
cag18_24     2.412758   2.191017   1.101 0.271192    
cag25_34     2.879946   2.228829   1.292 0.196743    
y81         -0.068626   0.017409  -3.942 8.91e-05 ***
y82         -0.040773   0.020196  -2.019 0.043894 *  
y83         -0.042178   0.019804  -2.130 0.033550 *  
y84         -0.013660   0.022350  -0.611 0.541290    
y85          0.009404   0.020782   0.453 0.651037    
y86          0.044095   0.022736   1.939 0.052854 .  
y87         -0.023960   0.021713  -1.103 0.270212    
y88          0.034758   0.021613   1.608 0.108253    
y89          0.025357   0.021867   1.160 0.246603    
y90          0.087170   0.021185   4.115 4.35e-05 ***
y91          0.038884   0.021359   1.821 0.069111 .  
y92          0.008150   0.022659   0.360 0.719183    
y93          0.008714   0.024004   0.363 0.716698    
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07893 on 690 degrees of freedom
Multiple R-squared:  0.2311,	Adjusted R-squared:  0.2055 
F-statistic: 9.019 on 23 and 690 DF,  p-value: < 2.2e-16
```

First stage equation results:

```
Residuals:
       Min         1Q     Median         3Q        Max
-0.3863488 -0.0567073 -0.0002044  0.0509309  0.4133336

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.014838   0.027520   0.539  0.58995
gpris       -1.031956   0.369963  -2.789  0.00543 **
gincpc       0.910199   0.214327   4.247 2.47e-05 ***
gpolpc       0.035315   0.067499   0.523  0.60101
cag0_14      3.379384   2.634893   1.283  0.20008
cag15_17     3.549945   5.766302   0.616  0.53834
cag18_24     3.358348   2.680839   1.253  0.21073
cag25_34     2.319993   2.706345   0.857  0.39161
cunem        0.523696   0.478563   1.094  0.27420
cblack      -0.015848   0.040104  -0.395  0.69285
cmetro      -0.591517   1.298252  -0.456  0.64880
y81         -0.056073   0.021735  -2.580  0.01009 *
y82          0.028462   0.038477   0.740  0.45973
y83          0.024703   0.037397   0.661  0.50911
y84          0.012870   0.029334   0.439  0.66098
y85          0.035403   0.027502   1.287  0.19844
y86          0.092186   0.034388   2.681  0.00752 **
y87          0.004771   0.029015   0.164  0.86944
y88          0.053271   0.027322   1.950  0.05161 .
y89          0.043086   0.027520   1.566  0.11790
y90          0.144265   0.035462   4.068 5.29e-05 ***
y91          0.061848   0.027650   2.237  0.02562 *
y92          0.026657   0.028533   0.934  0.35050
y93          0.022274   0.029610   0.752  0.45216
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.09547 on 690 degrees of freedom
Multiple R-Squared: -0.1246,    Adjusted R-squared: -0.1621
Wald test: 6.075 on 23 and 690 DF,  p-value: < 2.2e-16
```

2sls Model results:

```
Residuals:
      Min        1Q    Median        3Q       Max
-0.260197 -0.035356  0.001369  0.035515  0.281333

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.027201   0.017048   1.596 0.111038
final1      -0.077488   0.025956  -2.985 0.002932 **
final2      -0.052956   0.018408  -2.877 0.004141 **
gincpc       0.209552   0.131317   1.596 0.110998
gpolpc      -0.028692   0.044006  -0.652 0.514613
cag0_14      2.617307   1.582611   1.654 0.098626 .
cag15_17    -1.608738   3.755564  -0.428 0.668522
cag18_24     0.953368   1.731188   0.551 0.582017
cag25_34    -1.031684   1.763248  -0.585 0.558669
cunem        0.161659   0.311169   0.520 0.603563
cblack      -0.004476   0.026212  -0.171 0.864451
cmetro      -1.418389   0.786043  -1.804 0.071595 .
y81          0.012411   0.013763   0.902 0.367484
y82          0.077350   0.015692   4.929 1.04e-06 ***
y83          0.076779   0.015393   4.988 7.73e-07 ***
y84          0.028976   0.017650   1.642 0.101113
y85          0.027905   0.016418   1.700 0.089639 .
y86          0.054149   0.017930   3.020 0.002622 **
y87          0.031272   0.017132   1.825 0.068378 .
y88          0.019245   0.017073   1.127 0.260030
y89          0.018465   0.017287   1.068 0.285820
y90          0.063593   0.016577   3.836 0.000136 ***
y91          0.026372   0.016891   1.561 0.118919
y92          0.019048   0.017937   1.062 0.288639
y93          0.013411   0.018976   0.707 0.479965
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.06237 on 689 degrees of freedom
Multiple R-squared:  0.1522,    Adjusted R-squared:  0.1226
F-statistic: 5.153 on 24 and 689 DF,  p-value: 6.032e-14
```

Robust Standard Errors:

```
t test of coefficients:

             Estimate Std. Error t value  Pr(>|t|)
(Intercept)  0.014838   0.037352  0.3972  0.691310
gpris       -1.031956   0.337115 -3.0611  0.002291 **
gincpc       0.910199   0.331371  2.7468  0.006175 **
gpolpc       0.035315   0.061312  0.5760  0.564812
cag0_14      3.379384   2.989355  1.1305  0.258670
cag15_17     3.549945   6.012030  0.5905  0.555067
cag18_24     3.358348   3.378616  0.9940  0.320571
cag25_34     2.319993   2.922700  0.7938  0.427594
cunem        0.523696   0.489223  1.0705  0.284785
cblack      -0.015848   0.041042 -0.3861  0.699520
cmetro      -0.591517   1.631600 -0.3625  0.717061
y81         -0.056073   0.025567 -2.1931  0.028630 *
y82          0.028462   0.039690  0.7171  0.473557
y83          0.024703   0.039525  0.6250  0.532177
y84          0.012870   0.032611  0.3947  0.693213
y85          0.035403   0.031332  1.1299  0.258899
y86          0.092186   0.039240  2.3493  0.019089 *
y87          0.004771   0.032615  0.1463  0.883743
y88          0.053271   0.028916  1.8423  0.065863 .
y89          0.043086   0.029786  1.4465  0.148478
y90          0.144265   0.036652  3.9361 9.123e-05 ***
y91          0.061848   0.030891  2.0021  0.045663 *
y92          0.026657   0.030598  0.8712  0.383943
y93          0.022274   0.034637  0.6431  0.520396
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

<u>Comment 6</u>***: You need to discuss what these instruments mean and why you think that they are valid.***

Introduction to the Model:

The primary goal of the model was to understand how changes in litigation outcomes affect the growth in prison populations across states. Using "final1" and "final2" as instrumental variables in a 2SLS framework, we sought to address potential endogeneity issues that could arise from reverse causality or omitted variable bias in the regression model.

Instrumental Variables Justification:

Relevance: "final1" (final decision on litigation in the current year) and "final2" (decision on litigation in the previous two years) were chosen based on their relevance to the judicial processes affecting prison populations. These variables are theoretically linked to the prison population growth as they reflect the legal system’s disposition of cases, which can lead to either an increase or decrease in the number of incarcerations.

Exogeneity: These instruments are assumed to be exogenous with respect to the error term in the main regression equation. This assumption is based on the notion that while legal outcomes are influenced by judicial policies and efficiency, they are unlikely to be directly affected by the crime rate itself, thus providing a valid tool for isolating the causal impact of litigation outcomes on prison population dynamics.

Results from the 2SLS Estimation:

Significant Negative Coefficients: The coefficients of "final1" and "final2" were found to be -0.0775 and -0.0530, respectively, both statistically significant with p-values < 0.01. This indicates that increases in final litigation decisions are associated with a reduction in the growth rate of the prison population, supporting the theory that effective and conclusive litigation acts as a deterrent, influencing the prison population by potentially reducing crime through increased perceived risks of punishment.

Model Fit and Diagnostic Tests:

Residual Standard Error: The model displayed a residual standard error of 0.06237 on 689 degrees of freedom, indicating a good fit.

R-squared Values: The model achieved a multiple R-squared of 0.1522 and an adjusted R-squared of 0.1226, demonstrating that while the model captures a significant portion of the variability in prison population growth, other factors not included in the model may still play a role.

F-Statistic: The F-statistic of 5.153 with a p-value of 6.032e-14 strongly suggests that the model as a whole is statistically significant, reinforcing the appropriateness of the instrument selection and model specification.

Conclusion on Endogeneity:

The application of the 2SLS methodology in this context effectively addresses the endogeneity concerns, providing a more reliable estimate of the causal effects of imprisonment rates on the prison population. By using "final1" and "final2" as instruments, the model isolates the influence of external judicial decisions from other factors, thereby offering a robust analysis of how changes in litigation outcomes impact the growth of the prison population.

OVERVIEW OF PROJECT:

Step 1: OLS Regression Analysis

Objective : Estimate the relationship between crime growth rate (`gcriv`) and several predictors including prison population growth rate (`gpris`), political and economic factors, and demographic shifts.

Method: You employed an Ordinary Least Squares (OLS) regression model, where `gcriv` is modelled as a function of `gpris`, economic factors like per capita income growth (`gincpc`) and police per capita growth (`gpolpc`), demographic changes (`cag0_14`, `cag15_17`, `cag18_24`, `cag25_34`), and other covariates (`cunem`, `cblack`, `cmetro`). Additionally, year dummies from 1981 to 1993 were includedto control for year-specific effects.

Result: The regression output shows significant coefficients for `gpris` and `gincpc`, indicating strong relationships with `gcriv`. The overall model fit as measured by R-squared is moderately strong, suggesting that the model explains a significant proportion of the variance in `gcriv`.

Step 2: **First stage Equation for `gpris`**

Objective: Examine how external legal factors (`final1` and `final2`) influence the growth rate of the prison population (`gpris`).

Method: A reduced form OLS model where `gpris` is regressed on instrumental variables `final1` and `final2` (indicating final decisions on litigation), along with the same set of economic, demographic, and other covariates used in Step 1.

Result: Both instruments (`final1` and `final2`) were statistically significant with negative coefficients, supporting their relevance as instruments. This model also provided an F-statistic for the joint significance of the instruments.

Step 3: Two-Stage Least Squares (2SLS) Regression

Objective: Address potential endogeneity of `gpris` in the `gcriv` model by using `final1` and `final2` as instruments.

Method: A 2SLS regression where `gpris` is instrumented with `final1` and `final2` in the first stage, and the second stage involves regressing `gcriv` on the predicted values of `gpris` from the first stage, along with other covariates.

Result: The 2SLS estimates showed a significant negative coefficient for `gpris`, indicating a robust inverse relationship between prison population growth and crime growth, adjusting for endogeneity. The standard error of this estimate was also examined for robustness against heteroskedasticity and serial correlation.

Conclusion

Our analyses comprehensively attempt to elucidate the complex relationships between legal decisions, economic and demographic factors, and their impacts on crime and prison populations. By employing OLS and 2SLS methodologies, we have not only adjusted for potential endogeneity but also strengthened the causal interpretations of our findings. This project clearly demonstrates the application of instrumental variables in analysing policy-relevant issues in criminal justice. The findings of our project suggest that policies increasing the rate of imprisonment could potentially deter crime, provided these are part of a broader strategy that includes addressing socioeconomic factors. The application of robust statistical techniques such as 2SLS and robust standard errors has ensured that the estimates obtained are reliable and reflective of true causal relationships, despite the complexities and potential data issues inherent in econometric analysis. In summary, our project successfully combines theoretical insights with empirical evidence to suggest that effective judicial processing and incarceration policies can be integral components of crime reduction strategies.