Fall 2022 Econometrics Project

Jessica Li

Professor Allin Cottrell

ECN215 Econometric Theory and Methods

02/12/2022

Introduction

Many have studied and found that the severe income inequality not only creates negative impacts on overall economic performance, but also on social and health problems for the society. Considering the extent of detriments it bring to sustaining the prosperity of human lives, United Nations lists reducing inequality within and Among countries as the United Nations sustainable Development Goals, specifically, the goal targets at "progressively achieve and sustain income growth of the bottom 40 percent of the population at a rate higher than the national average by 2030(UNDP, 2022)". This issue is also highlighted by Robert J.Shiller, the winner of the 2013 Economics Nobel Prize, who once said that the rising inequality in the United States and elsewhere is the most important problem(St. Louis Post-Dispatch, 2013).

Focusing on this problem, In the book, *The Spirit Level,* the authors Kate Pickett and Richard G.Wilkinson use statistical analysis to further indicate that the increasing income inequality, measured by Gini coefficient, intensifies a range of social problems internationally and across the states in the U.S. However, even though this project was an influential piece at the time it was published, it received some criticisms on their method and approaches, but the critics tend to agree on the general message that the book conveys. Therefore, based on the work done

by Pickett and Wilkinson and their criticism, with a narrower scope of study, my goal for the project is to assess whether the income inequality can be an account for a series of social outcomes in the United States, and if there are any other better explanatory variables for these problems, in order to paint a fuller picture of what factors drive American society the way it is, and if reducing inequality is the main way to resolve these problems.

Literature review

In Richard Wilkinson and Kate Pickett's work, *The Spirit Level,* they argue that higher income inequality tend to associate with less trust between people, lower life expectancy, higher infant mortality rates, higher obesity rates, worse mental illness, lower education quality, higher teenage birth rate, higher homicide rate, and higher imprisonment rates. They suggest that these are not merely impacts on the people in the lower income group, the effects run through the society as a whole. Though they mentioned the effects other factors could have on these health and social indices, their statistical analysis only presented and based on the linear regression between Gini coefficient and each indices. This is also a part of the reason that the book has been subject to many criticisms - critics have indicated that they did not take account of other causal factors which might affect the health and social outcome measures. Targeting on this specific issue, In 2015, Simone Rambotti from the University of Arizona proposed to look into the possible spuriousness of their analysis and assess the potential role of poverty and the interaction between poverty and inequality in terms of explaining the social outcomes internationally and nationally for U.S.

In his paper, *Recalibrating the Spirit Level*, Rambotti focused on life expectancy and index of social and health problems for his research, and reached to the conclusion that poverty

has a significant adverse effect on both outcomes, and the effect of inequality becomes nonsignificant when poverty is taken into account. Therefore, He hence concluded that inequality and poverty are two different social issues that might have been confounded in the research done by Wilkinson and Pickett.

Inspired by Rambotti's work, I thereby decided to introduce two additional economic measures as the additional potential explanatory variables to the model: poverty rate and education scores, and take the possible influences from other social outcomes into account. Instead of focusing on two specific outcomes, I will apply these additional variables across the social outcomes mentioned in Wilkinson and Pickett's work, in order to build a better model that explains the social and health problems that exist in the United States.

The Data

Since my project is based on *The Spirit Level*, I decided to use the same dataset collected by Wilkson and Pickett with one additional economic indicator as the potential alternative explanatory variable as opposed to the income inequality: the poverty rate by state, which is suggested by Simone Rambotti in his study. A detailed account for the meaning of each variable used in this project and their source is organized in a table below:

Variable Name	Description	Source
State	The name of 50 states in the United States	
Gini	The measure of income inequality (varies between 0 to 1. 0 represents the perfect inequality and 1 represents the perfect inequality) by	U.S. Census Bureau, Table S4. Gini Ratios by State: 1969, 1979, 1989, 1999

	states in 1999.	
Poverty Rate	Number of people in poverty per a 1,000 people by states in 1999	U.S. Census Bureau
Life Expectancy	Life expectancy at birth for men and women by states in 2000	US Census Bureau. Population Division, Interim State Population Projections, Table 2. Internet release date: April 21, 2005.
Infant Mortality	Deaths in the first year of life per 1, 000 live births by states in 2002	US National Center for Health Statistics. Table 105, Statistical abstract of the United States. Washington, DC: CDC, 2006.
Mental Illness	Average number of days in the past month when mental health was not good by states in 2000	H. S. Zahran, R. Kobau, D. G. Moriarty, M. M. Zack, J. Holt, R. Donehoo. 'Health-related quality of life surveillance - United States, 1993-2002'. MMWR Surveill Summ 2005; 54(4): 1-35.
Education scores	Combined average of math and reading scores for 8th graders by states in 2003	US Department of Education NCfES. The Nation's Report Card: Reading Highlights 2003. Washington, DC, 2004. US Department of Education NCfES. The Nation's Report Card: Mathematics Highlights 2003. Washington, DC, 2004.
Teenage Birth Rate	Birth per 1,000 women aged 15-19 years by states in 2000.	US Census Bureau. Statistical Abstract of the United States: 2000 (120th Edition). Washington: Census Bureau, 2000.
Homicides	Homicide rate per 100,000 by	Federal Bureau of

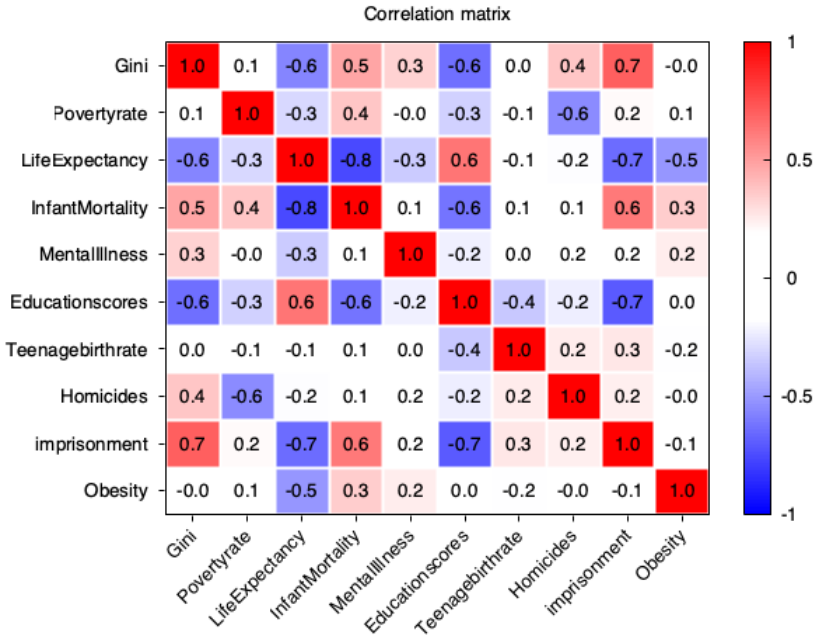
	states in 1999	Investigation. Crime in the United States 1999.
Imprisonment	Prisoners per 100,000 by states in 1998	US Department of Justice BoTS. Incarceration rates for prisoners under State or Federal jurisdiction.
Obesity	Percentage of the population with BMI >30, averaged for men and women by states in 1999	M. Ezzati, H. Martin, S. Skjold, S. Vander Hoorn, C. J. Murray. Trends in national and state-level obesity in the USA after correction for self-report bias: analysis of health surveys'. J R Soc Med 2006; 99(5): 250-7

For the purpose of this project, I group the above variables into two categories: the Economic indicators (Gini, poverty rate, and education scores) and the social outcome measures (Life expectancy, infant mortality, mental illness, teenage birth rate, homicides, imprisonment and obesity).

Modeling

There are two stages of my modeling for the project that serve different purposes. In the first stage, the dependent variables will be each social outcome measures and the regressors will only be selected from economic measures, in order to examine the impact of economic activity on the social problems and if the alternative variables can rule out the explanatory power of income inequality, measured by Gini ratio. In the second stage, I will expand the range of regressors to be both economic and social indicators, in order to better explain what factors contribute towards certain social problems to what extent in the society as a whole.

For both stages, I examined the correlation matrix of each variable against each other and used it as a reference for selecting independent variables for each dependent variable. It's noteworthy that there is a high correlation between Life expectancy and infant mortality that might incur a multicollinearity problem, but the later model suggests that there isn't a case where these two variables come to be the independent variables at the same time, so this problem is avoided.



First Stage (see detail models in Appendix)

For each social outcomes, I compared the adjusted R-Squared values between the model use solely Gini as the dependent variable and the model that selected dependent variables based on the correlation matrix, and the all appear to have a substantial improvement expect for the mental illness(from `0.089079 to 0.070758`), suggesting that the Gini coefficient can provide the strongest explanation for this problem by itself.

However, except for Teenage birth rate and Obesity, Gini ratio still remains significant in the new models with the introduction of other two economic indicators. This suggests that looking at other factors do help paint a clearer picture of what contributes to certain social problems, but these factors do not necessarily invalidate the explanatory power of Gini ratio as Rambotti suggests.

I want to highlight that for the model for the social outcome Obesity, the adjusted R square value remain negative before and after introducing new economic indicators as dependent variables(**-0.019319 to -0.009663**), which means that this model only have very weak explanatory power, so I decide not proceed this one further for testing because it's unnecessary and leave it to the stage two.

Afterwards, I have applied three tests to the preliminary models, which will be further explained in the testing section. Majority of them stay the same with the initial model, except that I have adjusted the nonlinearity for Teenage birth rate, and nonlinearity and heteroskedasticity for Imprisonment.

Second Stage (see detail models in Appendix):

In this stage, I build models with the dependent variables suggested by the correlation matrix against each social outcome. Similar to the first stage, I later adjusted the model after running four hypothesis testing, and the majority of them achieved significant improvement in the adjusted R square value. In addition to the Teen Birth Rate and mental illness, the adjusted R values for the rest of the models reached 0.47 and above with the highest being 0.934826 for the Infant Mortality model.

For the Teenage Birth Rate Model, the adjusted R squared value was in fact lower than the models in the first stage, this means that this particular aspect of social problem is not as susceptible to the other selected social indicators. However, considering the adjusted R squared value for the first stage model wasn't very high, it is possible that there are better predictors out there for this social outcome.

As for the mental illness model, regardless of the subtle increase in the adjusted R squared, none of the dependent variables (Gini, LifeExpectancy, Educationscores, Obesity) appear to be significant by their large p-values, and the nonlinear test doesn't suggest any quadratic transformation that can improve the model. Similar to the Teenage birth rate, it is possible that there could be better predictors than the ones listed in my dataset. This could make sense because there are a number of scholars who regard genetics to be the main contributor to mental illness rather than socio-economic factors.

Testing (see details in appendix)

There are three hypotheses that I used for the models in the first stage, and I will be explaining their meaning and application in the model respectively.

1. Test for nonlinearity

By running the test for nonlinearity, I aim to determine whether there are any squares of the dependent variable that can be useful for improving the current model. The null hypothesis of the parameter is not significant is rejected at the 5% significance level, so if the p-value of any squared term is less than 0.05, then I will add it to the model to include one more significant predictor.

2. White's test for heteroskedasticity

White's test is used to test whether the variance of the error term in the regression model is constant, with its null hypothesis being the variance for every error term is equal. If the test statistic shows that the null hypothesis is rejected at the 5% level of significance, it means that the model is subject to the problem of heteroscedasticity, meaning that the variances are homogenous, and therefore violating the IID condition for the OLS model. In order to solve this problem, I will switch to the robust standard deviation if the white's test is rejected.

3. Joint F-test

I use F-test to test the joint significance of the dependent variables in the model. The null hypothesis for this test is that all the variables have no predictive power for the independent variable (coefficient for each variable equals to 0). If the test statistics fail to reject the null hypothesis at 5% level of significance, I will remove the insignificant regressors to improve the model.

For the models in the first stage, based on the test results, I added the square of poverty rate as an additional variable to the model whose dependent variable is teenage birth rate, and the square of Gini ratio for the imprisonment model. Based on the white's test, the p-value for the test statistic for the imprisonment model is 0.036693, which rejects the null hypothesis for being less than 0.05, so I switched to the robust standard error for this model.

For the second stage, considering the larger scale of the number of dependent variables, I applied one more hypothesis testing: the sum of coefficients as another joint test. This test calculates the sum of coefficients of selected variables, and the p-value alongside with the standard error. As the null hypothesis for this test is that the sum equals to 0, I will remove the insignificant regression if the p-value exceeds 0.05.

Based on the result, I added the square of Life Expectancy to the infant mortality model, the square of education scores and teenage birth rate to the imprisonment model, and square of Gini to the Obesity model. According to the joint coefficient sum test, there are two models that fail to reject the null hypothesis: imprisonment and obesity, respectively with the p-value of 0.748309 and 0.940321. Therefore, I removed the variables in the test that has shown to be no predictive power for these two models.

Conclusion

Based on the result of the econometric modeling and testing, there are two general conclusions that can be drawn: (i) For most social outcomes, the income inequality proves to be a significant dependent variable that accounts for dependent variables. (ii) Social outcomes are better explained by the impacts from both economic and social factors.

In answering the question of this project, I can conclude that higher income inequality can explain for lower life expectancy, higher mental illness, higher homicide rates, and higher imprisonment rates. However, there can be cases that other variables have stronger explanatory power to the negative social outcomes. In the case of infant mortality, it is possible that the poverty rate, which is the only one out of three economic indicators that have positive coefficient, serve as the main driver towards the problem rather than the income inequality. because it is shown that higher income inequality can lead to lower infant mortality and lower obesity. Therefore, even though most of the results confirm the argument made in *The Spirit level,* it also validates some points made in the book's criticism: As the income inequality is able to explain the worsening of some social problems, we should strive for a more egalitarian society for people to thrive. However, we can't reach to the conclusion that reducing the income equality

can *promised* a society with a better living conditions, because there are many more factors that can play a role in here: the relevant social problems, education level, as well as the poverty, which, as Rombitti suggests, is a separate issue apart from inequality and can be confounded with one another sometimes.

Suggestions for further work

I can see many ways that this project can be developed further. Since I based my project on Wilkinson and Pickett's work, so I didn't expand my dataset, and it appear that some variables (Teenage birth rate and the mental illness) were failed to be explained by the variables in the dataset, so I think exploring more reasonable economic and social indicators that can be used for explanatory variables would certainly be helpful in advancing my current models and the results. Besides, due to the limit in time and length, there are some aspects that I wasn't able to dig into: 1) the possible effects of the interaction terms between economic indicators, as what have been done in Rambotti's work and 2) determining the which predictor is more important among all the current predictors. In addition, I think it would also be interesting to look internationally, and to investigate the differences and similarities for what can be accounted for the social problems in each country.

References:

1. *Guardian News and Media. (2010, July 8). The spirit level is not on the level | Natalie Evans. The Guardian. Retrieved December 3, 2022, from https://www.theguardian.com/commentisfree/2010/jul/08/spirit-level-book-critique*
2. *Pickett, Kate, and Richard Wilkinson. The Spirit Level. Penguin Books, 2010.*

3. *Rambotti S. Recalibrating the spirit level: An analysis of the interaction of income inequality and poverty and its effect on health. Soc Sci Med. 2015 Aug;139:123-31. doi: 10.1016/j.socscimed.2015.02.026. Epub 2015 Feb 18. PMID: 25726520.*

Appendix:

Preliminary models for the first stage and the tests:

(the hypothesis testing is showing for the model 2 as an example for saving space)

```
Model 1: OLS, using observations 1-51
Dependent variable: LifeExpectancy

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const        92.1784      3.12092     29.54     7.04e-33 ***
  Gini        -34.0085      6.95281     -4.891    1.12e-05 ***

Mean dependent var   76.93725   S.D. dependent var   1.518283
Sum squared resid    77.44527   S.E. of regression   1.257186
R-squared            0.328077   Adjusted R-squared   0.314365
F(1, 49)             23.92507   P-value(F)           0.000011
Log-likelihood      -83.01839   Akaike criterion     170.0368
Schwarz criterion    173.9004   Hannan-Quinn         171.5132


Model 2: OLS, using observations 1-51
Dependent variable: LifeExpectancy

                   coefficient   std. error   t-ratio   p-value
  ------------------------------------------------------------
  const              67.6696      11.2813       5.998    2.71e-07 ***
  Gini              -19.6403       8.47374     -2.318    0.0249   **
  Povertyrate        -0.101721     0.0705242   -1.442    0.1558
  Educationscores     0.0711133    0.0300756    2.364    0.0222   **

Mean dependent var   76.93725   S.D. dependent var   1.518283
Sum squared resid    61.80605   S.E. of regression   1.146744
R-squared            0.463765   Adjusted R-squared   0.429537
```

F(3, 47)	13.54937	P-value(F)	1.69e-06
Log-likelihood	-77.26634	Akaike criterion	162.5327
Schwarz criterion	170.2600	Hannan-Quinn	165.4855

Hypothesis Testing:

Auxiliary regression for non-linearity test (squared terms):

  Test statistic: TR^2 = 6.2056,
  with p-value = P(Chi-square(3) > 6.2056) = 0.102025

White's test for heteroskedasticity:

Test statistic: TR^2 = 7.785655,
with p-value = P(Chi-square(2) > 7.785655) = 0.020388

Test on Model 2:

  Null hypothesis: the regression parameters are zero for the variables
    Gini, Povertyrate, Educationscores
  Test statistic: F(3, 47) = 13.5494, p-value 1.68942e-06

Model 3: OLS, using observations 1-51
Dependent variable: InfantMortality

	coefficient	std. error	t-ratio	p-value	
const	-5.05609	3.28097	-1.541	0.1297	
Gini	27.3084	7.30937	3.736	0.0005	***

Mean dependent var	7.182353	S.D. dependent var	1.483065
Sum squared resid	85.59205	S.E. of regression	1.321657
R-squared	0.221707	**Adjusted R-squared**	**0.205824**
F(1, 49)	13.95832	P-value(F)	0.000488
Log-likelihood	-85.56892	Akaike criterion	175.1378
Schwarz criterion	179.0015	Hannan-Quinn	176.6143

Model 4: OLS, using observations 1-51
Dependent variable: InfantMortality

	coefficient	std. error	t-ratio	p-value	
const	25.5594	11.2923	2.263	0.0283	**
Gini	9.42174	8.48203	1.111	0.2723	
Povertyrate	0.121488	0.0705931	1.721	0.0918	*

```
  Educationscores   -0.0887105     0.0301050   -2.947    0.0050  ***

Mean dependent var   7.182353   S.D. dependent var   1.483065
Sum squared resid    61.92699   S.E. of regression   1.147866
R-squared            0.436895   Adjusted R-squared   0.400952
F(3, 47)             12.15526   P-value(F)           5.19e-06
Log-likelihood      -77.31619   Akaike criterion     162.6324
Schwarz criterion    170.3597   Hannan-Quinn         165.585
```

```
Model 5: OLS, using observations 1-51
Dependent variable: MentalIllness

            coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       0.611690      1.04720     0.5841    0.5618
  Gini        5.66170       2.33296     2.427     0.0190  **

Mean dependent var   3.149020   S.D. dependent var   0.441983
Sum squared resid    8.719426   S.E. of regression   0.421838
R-squared            0.107298   Adjusted R-squared   0.089079
F(1, 49)             5.889520   P-value(F)           0.018955
Log-likelihood      -27.32592   Akaike criterion     58.65185
Schwarz criterion    62.51550   Hannan-Quinn         60.12826
```

```
Model 6: OLS, using observations 1-51
Dependent variable: MentalIllness

                   coefficient   std. error   t-ratio   p-value
  ------------------------------------------------------------
  const              1.29029      3.83477      0.3365   0.7380
  Gini               5.29494      3.08558      1.716    0.0926  *
  Educationscores   -0.00190753   0.0103613   -0.1841   0.8547

Mean dependent var   3.149020   S.D. dependent var   0.441983
Sum squared resid    8.713273   S.E. of regression   0.426059
R-squared            0.107928   Adjusted R-squared   0.070758
F(2, 48)             2.903646   P-value(F)           0.064506
Log-likelihood      -27.30792   Akaike criterion     60.61585
Schwarz criterion    66.41132   Hannan-Quinn         62.83047
```

Model 7: OLS, using observations 1-51
Dependent variable: Teenagebirthrate

	coefficient	std. error	t-ratio	p-value	
const	12.9278	4.32834	2.987	0.0044	***
Gini	2.95694	9.64272	0.3067	0.7604	

Mean dependent var	14.25294	S.D. dependent var	1.727698
Sum squared resid	148.9612	S.E. of regression	1.743567
R-squared	0.001915	**Adjusted R-squared**	**-0.018454**
F(1, 49)	0.094034	P-value(F)	0.760409
Log-likelihood	-99.69830	Akaike criterion	203.3966
Schwarz criterion	207.2603	Hannan-Quinn	204.8730

Model 8: OLS, using observations 1-51
Dependent variable: Teenagebirthrate

	coefficient	std. error	t-ratio	p-value	
const	44.1802	8.75272	5.048	6.85e-06	***
Povertyrate	-0.198185	0.0949435	-2.087	0.0422	**
Educationscores	-0.103053	0.0309964	-3.325	0.0017	***

Mean dependent var	14.25294	S.D. dependent var	1.727698
Sum squared resid	119.1001	S.E. of regression	1.575199
R-squared	0.201993	**Adjusted R-squared**	**0.168743**
F(2, 48)	6.074940	P-value(F)	0.004448
Log-likelihood	-93.99341	Akaike criterion	193.9868
Schwarz criterion	199.7823	Hannan-Quinn	196.2014

Model 9: OLS, using observations 1-51
Dependent variable: Homicides

	coefficient	std. error	t-ratio	p-value	
const	-2426.42	1027.65	-2.361	0.0222	**
Gini	6305.19	2289.41	2.754	0.0082	***

Mean dependent var	399.2941	S.D. dependent var	440.3810
Sum squared resid	8396978	S.E. of regression	413.9649
R-squared	0.134044	**Adjusted R-squared**	**0.116371**
F(1, 49)	7.584861	P-value(F)	0.008236
Log-likelihood	-378.6606	Akaike criterion	761.3211
Schwarz criterion	765.1848	Hannan-Quinn	762.7975

Model 10: OLS, using observations 1-51
Dependent variable: Homicides

	coefficient	std. error	t-ratio	p-value	
const	5338.55	3075.90	1.736	0.0892	*
Gini	3431.44	2310.41	1.485	0.1442	
Povertyrate	−119.920	19.2288	−6.236	1.18e-07	***
Educationscores	−19.2098	8.20025	−2.343	0.0234	**

Mean dependent var	399.2941	S.D. dependent var	440.3810
Sum squared resid	4594712	S.E. of regression	312.6657
R-squared	0.526161	**Adjusted R-squared**	**0.495916**
F(3, 47)	17.39658	P-value(F)	9.78e-08
Log-likelihood	−363.2849	Akaike criterion	734.5699
Schwarz criterion	742.2972	Hannan-Quinn	737.5227

Model 11: OLS, using observations 1-51
Dependent variable: imprisonment

	coefficient	std. error	t-ratio	p-value	
const	−2742.43	465.531	−5.891	3.44e-07	***
Gini	7002.49	1037.11	6.752	1.61e-08	***

Mean dependent var	395.7843	S.D. dependent var	257.9282
Sum squared resid	1723165	S.E. of regression	187.5277
R-squared	0.481965	**Adjusted R-squared**	**0.471393**
F(1, 49)	45.58822	P-value(F)	1.61e-08
Log-likelihood	−338.2760	Akaike criterion	680.5520
Schwarz criterion	684.4156	Hannan-Quinn	682.0284

Model 12: OLS, using observations 1-51
Dependent variable: imprisonment

	coefficient	std. error	t-ratio	p-value	
const	2670.92	1499.14	1.782	0.0811	*
Gini	4076.79	1206.26	3.380	0.0014	***
Educationscores	−15.2167	4.05057	−3.757	0.0005	***

Mean dependent var	395.7843	S.D. dependent var	257.9282
Sum squared resid	1331643	S.E. of regression	166.5610
R-squared	0.599668	**Adjusted R-squared**	**0.582988**

```
F(2, 48)              35.95029   P-value(F)           2.87e-10
Log-likelihood      -331.7034    Akaike criterion     669.4067
Schwarz criterion    675.2022    Hannan-Quinn         671.6214
```

```
Model 13: OLS, using observations 1-51
Dependent variable: Obesity

             coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const       36.5842        9.85895     3.711    0.0005  ***
  Gini        -5.02683      21.9639     -0.2289   0.8199

Mean dependent var   34.33137   S.D. dependent var   3.933624
Sum squared resid    772.8436   S.E. of regression   3.971438
R-squared            0.001068   Adjusted R-squared  -0.019319
F(1, 49)             0.052381   P-value(F)           0.819924
Log-likelihood      -141.6813   Akaike criterion     287.3625
Schwarz criterion    291.2262   Hannan-Quinn         288.8390
```

```
Model 14: OLS, using observations 1-51
Dependent variable: Obesity

               coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const          32.5734       2.49656     13.05     1.46e-17 ***
  Povertyrate     0.162360     0.224840     0.7221   0.4737

Mean dependent var   34.33137   S.D. dependent var   3.933624
Sum squared resid    765.5232   S.E. of regression   3.952584
R-squared            0.010530   Adjusted R-squared  -0.009663
F(1, 49)             0.521452   P-value(F)           0.473655
Log-likelihood      -141.4386   Akaike criterion     286.8772
Schwarz criterion    290.7408   Hannan-Quinn         288.3536
```

Final models for the first stage (include only the ones have changed)

```
Model 15: OLS, using observations 1-51
Dependent variable: Teenagebirthrate

                   coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------------
```

```
  const              56.2165      8.88516       6.327    8.60e-08 ***
  Povertyrate         2.20259     0.763373      2.885    0.0059   ***
  sq_Povertyrate     -0.122967    0.0388442    -3.166    0.0027   ***
  Educationscores    -0.187887    0.0390757    -4.808    1.60e-05 ***

Mean dependent var   14.25294   S.D. dependent var   1.727698
Sum squared resid    98.16874   S.E. of regression   1.445232
R-squared            0.342240   Adjusted R-squared   0.300255
F(3, 47)             8.151547   P-value(F)           0.000179
Log-likelihood      -89.06485   Akaike criterion     186.1297
Schwarz criterion    193.8570   Hannan-Quinn         189.0825


Model 16: OLS, using observations 1-51
Dependent variable: imprisonment
Heteroskedasticity-robust standard errors, variant HC1

                    coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------------
  const              20228.5       2854.95       7.085    6.07e-09 ***
  Gini              -78624.9      13526.8       -5.813    5.17e-07 ***
  sq_Gini            90325.1      15089.2        5.986    2.83e-07 ***
  Educationscores      -10.3701       3.63691   -2.851    0.0064   ***

Mean dependent var   395.7843   S.D. dependent var   257.9282
Sum squared resid    734303.5   S.E. of regression   124.9939
R-squared            0.779246   Adjusted R-squared   0.765156
F(3, 47)             55.18058   P-value(F)           1.96e-15
Log-likelihood      -316.5246   Akaike criterion     641.0492
Schwarz criterion    648.7765   Hannan-Quinn         644.0020
```

<u>Preliminary models for the second stage and the tests:</u>

(the hypothesis testing is showing for the model 16 as an example)

```
Model 16: OLS, using observations 1-51
Dependent variable: LifeExpectancy

                    coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------------
  const             84.4962       6.70040       12.61     4.85e-16 ***
  Gini              -5.84097      5.39121       -1.083    0.2847
  Povertyrate       -0.0388077    0.0402952     -0.9631   0.3409
  InfantMortality   -0.233492     0.103200      -2.263    0.0288   **
  MentalIllness     -0.383575     0.235867      -1.626    0.1112
```

```
  Educationscores    0.0199156    0.0197251      1.010    0.3183
  imprisonment      -0.00234416   0.000607185   -3.861    0.0004   ***
  Obesity           -0.177019     0.0294194     -6.017    3.45e-07 ***

Mean dependent var   76.93725   S.D. dependent var   1.518283
Sum squared resid    17.28722   S.E. of regression   0.634057
R-squared            0.850014   Adjusted R-squared   0.825598
F(7, 43)             34.81347   P-value(F)           1.03e-15
Log-likelihood      -44.77848   Akaike criterion     105.5570
Schwarz criterion    121.0116   Hannan-Quinn         111.4626
```

Hypothesis testing

```
Auxiliary regression for non-linearity test (squared terms)
OLS, using observations 1-51:

Test statistic: TR^2 = 10.2023,
with p-value = P(Chi-square(7) > 10.2023) = 0.177394


Joint Test on Model 16:

Null hypothesis: the regression parameters are zero for the variables
Gini, Povertyrate, InfantMortality, MentalIllness, Educationscores,
imprisonment, Obesity
Test statistic: F(7, 43) = 34.8135, p-value 1.02611e-15


Coefficient sum joint test on Model 16:

Variables: const Gini Povertyrate MentalIllness Educationscores
   Sum of coefficients = 78.2528
   Standard error = 5.80355
   t(43) = 13.4836 with p-value = 4.81032e-17

White's test for heteroskedasticity:

Test statistic: TR^2 = 36.554242,
with p-value = P(Chi-square(35) > 36.554242) = 0.396435



Model 17: OLS, using observations 1-51
Dependent variable: InfantMortality

                   coefficient   std. error    t-ratio   p-value
  ------------------------------------------------------------
  const             46.8214       19.2996        2.426    0.0194  **
  Gini              -3.79316       7.87307      -0.4818   0.6323
```

```
  Povertyrate         0.0672577     0.0589453     1.141    0.2600
  LifeExpectancy     -0.385082      0.210363     -1.831    0.0739  *
  Educationscores    -0.0430944     0.0283723    -1.519    0.1359
  imprisonment        0.00136499    0.00101409    1.346    0.1852
  Obesity             0.0593366     0.0578467     1.026    0.3106

Mean dependent var    7.182353   S.D. dependent var   1.483065
Sum squared resid     38.45161   S.E. of regression   0.934826
R-squared             0.650358   Adjusted R-squared   0.602679
F(6, 44)              13.64048   P-value(F)           1.11e-08
Log-likelihood       -65.16403   Akaike criterion     144.3281
Schwarz criterion     157.8508   Hannan-Quinn         149.4955


Model 18: OLS, using observations 1-51
Dependent variable: MentalIllness

                    coefficient    std. error   t-ratio    p-value
  ------------------------------------------------------------------
  const              1.70506       6.40510       0.2662    0.7913
  Gini               5.04757       3.33016       1.516     0.1364
  LifeExpectancy    -0.0207168     0.0750943    -0.2759    0.7839
  Educationscores   -0.000151649   0.0121454    -0.01249   0.9901
  Obesity            0.0237867     0.0217012     1.096     0.2787

Mean dependent var    3.149020   S.D. dependent var   0.441983
Sum squared resid     8.090059   S.E. of regression   0.419370
R-squared             0.171733   Adjusted R-squared   0.099710
F(4, 46)              2.384409   P-value(F)           0.064940
Log-likelihood       -25.41553   Akaike criterion     60.83106
Schwarz criterion     70.49019   Hannan-Quinn         64.52210


Model 19: OLS, using observations 1-51
Dependent variable: Teenagebirthrate

                   coefficient    std. error   t-ratio   p-value
  ---------------------------------------------------------------
  const            34.1165       12.0841        2.823    0.0070  ***
  Educationscores  -0.0745463     0.0434081    -1.717    0.0925  *
  Homicides         0.000666148   0.000543420   1.226    0.2264
  imprisonment     -8.37560e-05   0.00128041   -0.06541  0.9481

Mean dependent var   14.25294   S.D. dependent var   1.727698
Sum squared resid    125.8799   S.E. of regression   1.636550
R-squared            0.156567   Adjusted R-squared   0.102731
F(3, 47)             2.908216   P-value(F)           0.044265
Log-likelihood      -95.40518   Akaike criterion     198.8104
Schwarz criterion    206.5377   Hannan-Quinn         201.7632
```

```
Model 20: OLS, using observations 1-51
Dependent variable: Homicides

                    coefficient   std. error   t-ratio  p-value
  ------------------------------------------------------------
  const              4712.28       3917.56       1.203   0.2353
  Gini               3894.47       2830.26       1.376   0.1756
  Povertyrate        -117.385        21.1386    -5.553   1.43e-06 ***
  Educationscores     -18.2164       10.3994    -1.752   0.0866   *
  Teenagebirthrate      9.92810      31.5748     0.3144  0.7546
  imprisonment         -0.0455135     0.280648  -0.1622  0.8719

Mean dependent var   399.2941   S.D. dependent var   440.3810
Sum squared resid     4583379   S.E. of regression   319.1439
R-squared            0.527329   Adjusted R-squared   0.474811
F(5, 45)             10.04075   P-value(F)           1.68e-06
Log-likelihood      -363.2220   Akaike criterion     738.4439
Schwarz criterion    750.0349   Hannan-Quinn         742.8732


Model 21: OLS, using observations 1-51
Dependent variable: imprisonment

                    coefficient    std. error   t-ratio  p-value
  -------------------------------------------------------------
  const              2206.64       2797.09        0.7889  0.4344
  Gini               4028.64       1329.97        3.029   0.0041  ***
  LifeExpectancy      -29.2998       25.4399     -1.152   0.2557
  InfantMortality      26.5525       25.6730      1.034   0.3067
  Educationscores      -6.90703       5.05738    -1.366   0.1790
  Teenagebirthrate     22.3261       15.3626      1.453   0.1532
  Homicides            -0.0224756     0.0583640  -0.3851  0.7020

Mean dependent var   395.7843   S.D. dependent var   257.9282
Sum squared resid     1137319   S.E. of regression   160.7737
R-squared            0.658088   Adjusted R-squared   0.611463
F(6, 44)             14.11466   P-value(F)           6.97e-09
Log-likelihood      -327.6810   Akaike criterion     669.3620
Schwarz criterion    682.8848   Hannan-Quinn         674.5295


Model 22: OLS, using observations 1-51
Dependent variable: Obesity

                    coefficient   std. error   t-ratio   p-value
```

```
  ------------------------------------------------------------
  const             222.443       47.2118       4.712    2.30e-05 ***
  Gini              -79.1541      21.0498      -3.760    0.0005   ***
  LifeExpectancy     -2.02755      0.521819    -3.886    0.0003   ***
  InfantMortality    -0.0597241    0.488957    -0.1221   0.9033
  MentalIllness       1.20184      1.13842      1.056    0.2966

Mean dependent var   34.33137   S.D. dependent var   3.933624
Sum squared resid    428.1406   S.E. of regression   3.050804
R-squared            0.446611   Adjusted R-squared   0.398490
F(4, 46)             9.281030   P-value(F)           0.000014
Log-likelihood      -126.6203   Akaike criterion     263.2407
Schwarz criterion    272.8998   Hannan-Quinn         266.9317
```

Final models for the second stage (include only the ones have changed):

```
Model 23: OLS, using observations 1-51
Dependent variable: InfantMortality

                     coefficient     std. error   t-ratio  p-value
  ------------------------------------------------------------------
  const              1085.20         389.258        2.788   0.0079  ***
  Gini                 -8.63239        7.59525     -1.137   0.2620
  Povertyrate          -0.0249924      0.0651376   -0.3837  0.7031
  LifeExpectancy      -27.3154        10.0864      -2.708   0.0097  ***
  sq_LifeExpectancy     0.174565       0.0653688    2.670   0.0106  **
  Educationscores      -0.0304620      0.0269983   -1.128   0.2655
  imprisonment          0.000421958    0.00101356   0.4163  0.6793
  Obesity               0.0590660      0.0541939    1.090   0.2818

Mean dependent var   7.182353   S.D. dependent var   1.483065
Sum squared resid    32.98170   S.E. of regression   0.875795
R-squared            0.700096   Adjusted R-squared   0.651274
F(7, 43)             14.33987   P-value(F)           1.92e-09
Log-likelihood      -61.25111   Akaike criterion     138.5022
Schwarz criterion    153.9568   Hannan-Quinn         144.4079


Model 24: OLS, using observations 1-51
Dependent variable: imprisonment

                      coefficient    std. error   t-ratio  p-value
  ------------------------------------------------------------------
  const                54996.3       17870.3        3.078   0.0036  ***
  Gini                -60156.2       19171.5       -3.138   0.0031  ***
  sq_Gini              68566.6       21263.9        3.225   0.0024  ***
  LifeExpectancy         -51.7636       12.6161    -4.103   0.0002  ***
```

```
  Educationscores       −294.262       157.543     −1.868   0.0686  *
  sq_Educationscor~        0.547583      0.294930   1.857   0.0702  *
  Teenagebirthrate       244.160        73.4387     3.325   0.0018  ***
  sq_Teenagebirthr~       −7.04879       2.35205   −2.997   0.0045  ***

Mean dependent var   395.7843   S.D. dependent var   257.9282
Sum squared resid    417042.7   S.E. of regression   98.48181
R-squared            0.874624   Adjusted R-squared   0.854215
F(7, 43)             42.85281   P-value(F)           2.33e-17
Log-likelihood      −302.0984   Akaike criterion     620.1968
Schwarz criterion    635.6514   Hannan-Quinn         626.1024


Model 25: OLS, using observations 1-51
Dependent variable: Obesity

                  coefficient   std. error   t-ratio   p-value
  ------------------------------------------------------------
  Gini              925.723     107.366       8.622    2.57e-11 ***
  sq_Gini         −1082.46      129.024      −8.390    5.71e-11 ***
  LifeExpectancy     −2.11131     0.292277   −7.224    3.34e-09 ***

Mean dependent var   34.33137   S.D. dependent var   3.933624
Sum squared resid    359.7137   S.E. of regression   2.737524
Uncentered R-squared 0.994092   Centered R-squared   0.535055
F(3, 48)             2692.130   P-value(F)           1.83e-53
Log-likelihood      −122.1797   Akaike criterion     250.3594
Schwarz criterion    256.1548   Hannan-Quinn         252.5740
```