Income Inequality and Crime: Dynamic Impacts Using Panel Data in INDIA

Indian Institute of Technology, Kanpur

ECO342A, Econometrics II

Debashree Priya Sahoo

April 21,2023

Abstract:

Income inequality has long been recognized as a significant social and economic issue, with potential implications for various aspects of society, including crime rates. This research paper utilizes panel data set covering from year 2016-2021 across 35 Indian states and union territories analysis to look at the relationship between income inequality and crime rates using a comprehensive theoretical framework and empirical analysis of data from a range of sources. Using SPI Index as an indicator, I recorded an association between income inequality and crime while controlling for other relevant factors and employing various econometric techniques, including generalized method of moments (GMM), threshold modeling, panel unit root tests, and autoregressive distributed lag (ARDL) cointegration analysis. Our model shows a positive link between SPI index and income inequality suggesting the presence of some social theories that contributes to crime.

Introduction:

Income inequality has become a pressing issue in many countries, including India, as the gap between the rich and the poor continues to widen. Concerns have been expressed regarding this trend's possible effects, especially how it may affect crime rates. Many other factors, such as income levels, poverty rates, and exposure to a lack of human capital, have an impact on crime involvement. In the context of economic theory, criminal behavior is frequently viewed as a rational cost-benefit analysis in which people select the most profitable endeavor. The degree of financial disparity in a society can shape the perceived advantages and disadvantages of engaging in criminal activity, which is a crucial component that may influence this decision-making process. The connection between income inequality and crime is complex and multifaceted, and there is much debate among scholars about the nature and extent of this connection.

Endogeneity is a frequent problem in empirical research that can result from several factors, including measurement error, unobserved heterogeneity, and the assessment of many important variables at the same time. Endogeneity can be an important concern, particularly when researching the connection between crime and wealth inequality. For instance, shared underlying variables like economic, social, or cultural elements may jointly drive income disparity and criminality. Furthermore, both crime and income inequality may be impacted by unobservable variation, such as individual or geographical variables. As a result of these shared elements, the underlying link between crime and income inequality may be hidden, which might result in endogeneity.

To address these endogeneity issues, researchers may employ instrumental variables (IVs) and the generalized method of moments (GMM) estimation technique. These techniques can help to improve the validity and reliability of the empirical findings, providing more robust and credible estimates of the causal relationships between variables of interest. IVs are external variables that are correlated with the endogenous variable of interest but are not directly related to the outcome of interest. By including valid IVs in the model, researchers can account for the potential bias introduced by endogeneity and unobserved heterogeneity, helping to improve the internal validity of the findings. GMM is a statistical technique that uses moment conditions to estimate the parameters of a model. It can be used to address endogeneity by utilizing moment conditions based on the orthogonality conditions between the instruments and the errors in the model. GMM estimation can provide consistent estimates even in the presence of endogeneity, making it a useful method for tackling the potential issues arising from measurement error, unobserved heterogeneity, and simultaneous determination of variables.

Crime involvement has long been acknowledged in the economics community as a complicated issue driven by a variety of variables. Economic theories, such as the rational choice theory, posit that criminals engage in a cost-benefit analysis before committing a crime. The perceived advantages of committing a crime, like financial gain, must exceed the believed drawbacks, such the possibility of being detected and punished. This cost-benefit analysis can be considerably impacted by income disparity. In areas with high income inequality, the opportunity costs of engaging in crime may be perceived as relatively low, as the potential benefits of targeting wealthier individuals or areas may seem more attractive. On the other hand, in areas with low-income inequality, the perceived benefits of crime may be lower due to less wealth disparity, making it less attractive for potential criminals.

.

India offers a rich backdrop for researching the link between income inequality and crime rates because of its different states and range of income disparity. In addition to the main explanatory variable highlighted (SPI Index), this study also identifies the impact of control variables (population density, unemployment rate, gender ratio, and inflation rate (Consumer price index, CPI)), including two instruments (Gross State Domestic Product, GSDP and education attainment(net enrollment ratio)) on crime rates across Indian States. Panel data analysis can provide important insights into the dynamics of this connection by adding control variables and instruments and applying techniques such as threshold effects, panel unit root, GMM and ARDL cointegration, can offer valuable insights into the dynamics of this relationship. Policymakers and stakeholders may build strategies to address the underlying causes of crime and improve social and economic well-being for all residents by understanding the effect of income disparity on crime rates in Indian states.

Literature Review:

The association between income inequality and crime rates has gained significant attention from scholars in various fields such as criminology, economics, sociology, and policymaking. There are divergent views among researchers regarding this relationship. Some studies, such as those by **Fajnzylber et al. (2002)** found positive and significant relationships between income inequality and violent crime in a panel of countries. **Demombynes and Ozler (2005),** showed income inequality is associated with property and violent crime in South African neighborhoods. **Enamorado et al. (2016)** showed that a one-point increase in the Gini coefficient, a measure of income inequality, raises drug-related homicides. However, there are also studies, such as those by **Doyle et al. (1999) and Neumayer (2005)** that have found no significant link between income inequality and violent crime. A **time series regression by Saridakis (2004**) found a short-term but no long-run relationship between income inequality and crime. **Brush (2007)** found mixed evidence, with a positive association between inequality and crime across cross section of U.S. countries in some analyses and a negative association in time-series analysis. **Chintrakarn and Herzer (2012)** found a negative correlation between inequality and crime.

Nagasubramaniyan and Joseph (2022) found a negative linkage between the Gini ratio and violent crime rates in India using GMM and other data panel techniques. **Kelly (2000) and Choe (2008)** found that income inequality had no influence on property crime but had a significant and robust impact on multiple indicators of violent crime, based on state-level data for the United States. In addition to income inequality, other factors such as population density and education have also been explored in relation to crime**. Andresen (2011)** argued that densely populated areas face more crime than rural areas, and studies **by Atems (2020) and Akçomak and ter Weel (2012)** showed a significant positive impact of population density on crime. **Buonanno and Leonida (2006)** found that education significantly negatively impacts crime of all types using average years of schooling as a measure. Similarly, studies on the relationship between Gross Domestic Product (GDP) as an income variable and crime have also shown varying results. **Levitt (2002)** provided evidence of a significant positive impact of income on crime rates, while **Lin (2007)** demonstrated that the impact of GDP on crime can differ depending on the type of crime.

The mixed findings on the relationship between income inequality and crime may be attributed to endogeneity issues, including measurement error, unobserved heterogeneity, and simultaneous determination of crime and income inequality. Causality is difficult to establish as both income inequality and crime may be influenced by common factors. For instance, changes in crime rates may impact income inequality if higher-income individuals leave the area, while policies aimed at reducing crime may also decrease inequality. Additionally, cross-country studies are prone to measurement error and unobserved heterogeneity, further complicating comparisons. In summary, determining a clear cause-and-effect relationship between income inequality and crime is challenging due to these endogeneity concerns.

The base paper estimated a panel SVAR model to assess the relationships among multiple variables in a system and identify the relative importance of various shocks on the system. for US state-level data set from 1960-2015 for fifty U.S. states and D.C. (56 observations per state). More than 2800 observations were obtained by combining the cross-sectional and time-series aspects of the data. Different reference papers also show institutions' qualities influence how closely economic inequality and crime are related. Overall, the literature on the impact of income inequality on crime rates is complex, with varying findings from different studies in different contexts and there is a need for further research to better understand these dynamics.

Data

The purpose of this study is to examine the causal relationship between income inequality and the crime rate in India using a panel dataset from 2016 to 2021, covering 35 Indian states and union territories. Crime rates are measured by summing the Indian Penal Code (IPC) and Special and Local Laws (SLL) crime values obtained for each of these Indian states. IPC stands for the

Indian Penal Code. It is the main criminal code of India that defines various offenses and prescribes penalties for them. The IPC is a comprehensive legislation that covers a wide range of criminal offenses. For ex- crimes against persons (such as murder, assault, and kidnapping), crimes against property (such as theft, robbery, and burglary), crimes against public order (such as rioting and unlawful assembly), and crimes against the state (such as sedition and treason) and etc. SLL stands for Special and Local Laws. This category includes crimes that are classified under special or local laws enacted by the central or state governments in India, apart from the Indian Penal Code. Special laws are enacted to address specific issues or offenses that may require specialized legal provisions beyond what is covered by the IPC. Examples of special and local laws include the Narcotic Drugs and Psychotropic Substances Act, the Arms Act, the Unlawful Activities (Prevention) Act, and the Protection of Children from Sexual Offences Act, among others. This is my dependent variable here.

The Social Progress Index (SPI) is a composite index that measures the social and environmental well-being of countries. One of the indicators included in the SPI is the Gini coefficient, which is a measure of income inequality. The Gini coefficient ranges from 0 (perfect equality) to 1 (perfect inequality), with higher values indicating greater income inequality. It is an indirect indicator for income inequality. Some of the dimensions and indicators included in the SPI, such as access to basic education, health, and infrastructure, may be related to income inequality. It is my main explanatory variable here (the endogenous regressor).

Control variables are factors that are held constant or kept consistent to isolate the effects of the independent variable on the dependent, can more accurately assess the impact of income inequality on crime rates without the influence of other factors that might be related to both income inequality and crime rates. My 4 control variables include unemployment rate, inflation rate, population density and gender ratio. In research it has been shown that there is a correlation between gender imbalance and higher crime rates. High levels of unemployment may lead to higher crime rates as individuals may turn to illegal activities to make ends meet.
High levels of inflation can lead to higher levels of poverty and economic instability, which can increase the likelihood of criminal activity. Population density is often used as an indicator of urbanization, with higher population density indicating higher concentrations of people in a specific area. Population density can influence social dynamics, community interactions, and resource allocation, which may have implications for crime rates.

Education (net enrollment rate) in primary education and Gross State Domestic Product (GSDP) are chosen as instrumental variables (IVs) to address endogeneity. IVs are expected to be correlated with the endogenous regressor but uncorrelated with the error term. Education and GSDP are hypothesized to be associated with lower income inequality, as education improves human capital and GSDP reflects economic development and growth. However, unequal distribution of economic benefits may contribute to income inequality.

Research Gap

- The problem of simultaneity bias in regression of crime on income inequality- the two variables are related here.
- Omitted variables and measurement errors can lead to biased estimates.
- Inadequate attention to mediating factors- such as poverty, education, employment opportunities, etc.
- Limited attention to policy implications
- Inadequate consideration for reverse causality

Defined Variables

Label	**Variable Description**	**Measure**	**Type**
crimes	Total number of Crimes	Sum of IPC and SLL crimes	dependent
spi	Social Progress Index	Social Progress Index across 36 Indian states and UTs	Independent(endogenous)
pd	Population Density (in Lakhs)	refers to the number of people per hundred thousand	Independent variable (Control variable)
gr	Gender Ratio	the number of females per thousand males in a given population	Independent variable (Control variable)
ur	Unemployment Rate	% of total labor force who are unemployed and are looking for a paid job	Independent variable (Control variable)
ir	Inflation Rate (CPI)	rate at which the general level of prices for goods and services is rising	Independent variable (Control variable)
edu	net enrollment rate in primary education	% tage of primary-school-age children who are enrolled in primary school.	Instrument
gsdp	Gross State Domestic Product in lakhs	Total economic output of a state in India.	Instrument

Descriptive Statistics

```
sum crimes pd gr ur ir edu gsdp
```

Variable	Obs	Mean	Std. Dev.	Min	Max
crimes	210	156385.3	207847.5	50	1377681
pd	210	378.8843	473.8365	.7	2317
gr	210	946.2286	52.51416	775	1085
ur	210	83.80952	43.83246	12	286
ir	210	4.805	1.959424	.4	12.4
edu	210	88.47405	9.498586	61.35	100
gsdp	210	5.89e+07	6.50e+07	120836	2.73e+08

Base Model

$$CMRit = \alpha 0 + \beta 1 CMRit{-}1 + \beta 2 Eit + \beta 3 Population\_density\mathrm{it} + \beta 4 Uemit + \beta 5 INFit + +\beta 6 \mathrm{gr}it + \beta 7 \mathrm{edu}it + \beta 8 GSDPCit + \varepsilon i\mathrm{t}$$

$CMRit$- Crime rate in state i at time t
$CMRit-1$-Crime rate in state i at time t-1 (lagged dependent variable)
$IEit$- Income inequality (SPI Index) in state i at time t
$GSDPCit$- Gross State Domestic Product in state i at time t
$Uemit$- The unemployment rate in state i at time t
$INFit$- Inflation rate for region i at time t
$Population_density$it- Population density for region i at time t
eduit- Education level (net enrollment rate in primary education) in state i at time t
grit- Demographic factors (gender ratio) in state i at time t
εit- Error term

Hypothesis

1. Will examine the relationship between income inequality and crime rates in India, similar to the studies done in the United States, and claim that higher income inequality causes higher crime rates. By using threshold regression and various panel data analysis techniques, we can better identify the causal effect of income inequality on crime rates and control for endogeneity.
2. In India, institutional quality moderates the association between income inequality and crime: Based on the results of the study indicated in the base paper for US, this hypothesis would investigate whether India's institutions have an impact on the relationship between income inequality and crime.

3. Will examine the relationship between income inequality and crime in India for IPC+SLL crimes and check how much they get affected by other factors present. Compare SPI Index with different crime measures.
4. Will try to analyze different policies focusing on poverty reduction, education, and employment opportunities that must have been implemented and check for the impact of income inequality on crime across different states.

Methodology

1. **OlS Regression-** This model can be used to analyze the relationship between crime rate and other factors mentioned, and estimate the coefficients ($\beta1$, $\beta2$, $\beta3$, $\beta4$, $\beta5$, $\beta6$, $\beta7$, $\beta8$) to quantify the strength and direction of these relationships.

2. **2SLS Regression-** 2SLS is a method used to estimate the parameters of a linear regression model when there are endogenous variables by using instrumental variables (IVs) in a two stage process to address the endogeneity to obtain consistent and unbiased estimates of the parameters.

3. **Endogeneity Test-**

Approach 1

Comparing the coefficients obtained from OLS and 2SLS regression can provide insights into the potential impact of endogeneity on the estimated coefficients. As in OLS endogeneity is not addressed but by using instrumental variables in 2SLS regression, you are addressing potential endogeneity by using a two-stage approach. If the coefficients obtained from OLS and 2SLS regression are significantly different, it may indicate that endogeneity is present in the model and that the coefficients obtained from OLS regression may be biased. In such cases, using 2SLS regression with instrumental variables can be considered as a more appropriate approach to obtain unbiased estimates of the coefficients.

Approach 2

Checking for endogeneity, overidentification test and weak or strong instrument test after running 2sls regression and interpreting them

4. **GMM Estimation-** In contrast to OLS and 2SLS estimations, GMM provides reliable estimates. GMM provides reliable estimates. Lagged values for the dependent variable are incorporated into the GMM mode which helps in internally transforming the data. As a result, the endogeneity issues are resolved, and the valid estimates are generated utilizing a strict GMM process. GMM estimation is particularly useful when dealing with models that have more instruments than endogenous variables, which is known as an overidentified model. In such cases, 2SLS may not be feasible due to the "curse of dimensionality" or other issues, and GMM can provide a viable alternative. GMM estimation uses moment conditions, which are functions of the data and parameters, to

estimate the model parameters. These moment conditions are chosen to be orthogonal to the error term.

Using the Arellano-Bover GMM estimator, the general model form to be used is

$$Crime_{i,t} = \alpha_0 + \alpha_1 Crime_{i,t-1} + \alpha_2 Income\_Inequality_{i,t} + \beta_1\, control_{i,t} + \beta_2\, instruments_{i,t} + \varepsilon_{i,t}$$

Crime,t- the crime rate for individual i for time period t
Crimei,t-1- One period lagged operator(previous year crime rate)
Income_Inequalityi,t - the income inequality measure for individual i for time period t
α1 and α2 - Parameters
Controli,t- vector of control variables that may affect crime rates,
β - vector of parameters to be estimated for the control variables.
εi,t-error term that captures unobserved factors affecting crime rates.

The three main endogeneity sources—unobserved heterogeneity, simultaneity, and dynamic endogeneity—can be controlled for using the GMM model. Instruments should be chosen such that they are correlated with the endogenous regressors but not correlated or orthogonal to the errors. In GMM one minimizes instruments, where W is the weighting matrix and Z is a matrix of instruments and the moment or orthogonality condition is E(Z'u)=0

5. Instrumental variable regression (RE, FE, BE, FD)- Instrumental variable (IV) regression is a statistical method used to estimate causal relationships between variables in the presence of endogeneity, which, which occurs when explanatory variables and error term are correlated in a regression model in a regression model. By using instrumental variables, IV regression allows researchers to obtain consistent estimates of the parameters of interest, even when the standard ordinary least squares (OLS) estimates may be biased due to endogeneity. It is here estimated for 4 categories- random effects, fixed effects, between effects and first difference.

6. Threshold Regression- Threshold regression with dynamic panel data is a statistical technique used to model the relationship between variables in panel data, where the relationship may change depending on the values of certain threshold variables. It allows for the estimation of separate coefficients for different regimes or states of the threshold variable. Dynamic panel data refers to panel data that includes lagged values of variables, allowing for the consideration of time dynamics and potential endogeneity issues. This approach can provide insights into how different variables interact and influence each other in different states or regimes and can help uncover complex relationships that may not be captured by traditional linear regression methods.

7. Panel Unit Root tests- Panel unit root tests are statistical tests used to determine if a panel data series exhibits unit root behavior, which implies that the series is non-stationary and has a stochastic trend. Panel unit root tests are useful in panel data analysis as they help to assess the stationarity of variables, which is an important consideration in econometric modeling. If a variable is found to have a unit root, it suggests that the variable is non-stationary and may require differencing or other techniques to transform it into a stationary series. Commonly used panel unit root tests include the Levin-Lin-Chu (LLC) test, the Im-Pesaran-Shin (IPS) test, and the Fisher-type panel unit root tests.

8. ARDL Approach to Cointegration and Bounds Test- If panel unit root tests indicate that the variables of interest in a panel dataset are non-stationary and exhibit unit root behavior, one common approach to address this issue is to use the Autoregressive Distributed Lag (ARDL) approach to cointegration. Cointegration is a statistical concept that captures the long-run relationship between non-stationary variables. The ARDL approach involves estimating an autoregressive model that includes lagged levels of the non-stationary variables and their first-difference terms, along with other relevant variables. The lag order in the ARDL model is determined based on standard model selection criteria, such as Akaike Information Criterion (AIC) or Bayesian Information Criterion (BIC). The ARDL model allows for the estimation of short-run and long-run dynamics between the variables. One of the advantages of the ARDL approach is that it allows for the estimation of cointegrating relationships in panel data without the need for all variables to be stationary. Bounds test can be used for the existence of cointegration using, If cointegration is confirmed, then the estimated coefficients can be interpreted as the long-run equilibrium relationship between the variables.

Results and Interpretation

1.OLS Results

```
. xtreg crimes L.crimes spi pd edu gr gsdp ur ir

Random-effects GLS regression                   Number of obs      =       175
Group variable: States                          Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0493                                          min =         5
     between = 0.9799                                          avg =       5.0
     overall = 0.8427                                          max =         5

                                                Wald chi2(8)       =    889.09
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      crimes |
         L1. |   .7521923    .055103    13.65   0.000     .6441924    .8601922
             |
         spi |     528.65   970.9643     0.54   0.586    -1374.405    2431.705
          pd |   60.29692   26.18302     2.30   0.021     8.979139    111.6147
         edu |  -1042.287   758.1131    -1.37   0.169    -2528.161    443.5875
          gr |   34.40321   136.5195     0.25   0.801    -233.1701    301.9765
        gsdp |   .0003497   .0001854     1.89   0.059    -.0000138    .0007131
          ur |  -117.4186   169.5974    -0.69   0.489    -449.8234    214.9862
          ir |   4076.941    3351.42     1.22   0.224    -2491.721     10645.6
       _cons |   21089.51   132244.4     0.16   0.873    -238104.7    280283.8
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  74536.263
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Here, Lagged crime rate, pd(population density), and gsdp(gross state domestic product) are statistically significant and have an effect on crime rate in India

2. 2SLS Regression

```
. ivregress 2sls crimes L.crimes pd gr ur ir ( spi= gsdp edu)

Instrumental variables (2SLS) regression          Number of obs   =        175
                                                  Wald chi2(6)    =     840.80
                                                  Prob > chi2     =     0.0000
                                                  R-squared       =     0.8249
                                                  Root MSE        =      88921

------------------------------------------------------------------------------
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         spi |    4614.11   1991.492     2.32   0.021     710.8573    8517.362
             |
      crimes |
         L1. |   .7084765   .0699773    10.12   0.000     .5713235    .8456294
             |
          pd |    114.513   28.84269     3.97   0.000     57.98238    171.0436
          gr |  -104.3182    140.898    -0.74   0.459    -380.4733    171.8369
          ur |  -.8536322   177.5242    -0.00   0.996    -348.7946    347.0874
          ir |   3724.059     3451.2     1.08   0.281    -3040.169    10488.29
       _cons |  -169920.2   144628.4    -1.17   0.240    -453386.6    113546.2
------------------------------------------------------------------------------
Instrumented:  spi
Instruments:   L.crimes pd gr ur ir gsdp edu

.
```

Overall, the 2SLS regression model suggests that SPI and pd have a statistically significant relationship with crimes after accounting for potential endogeneity or omitted variable bias through the use of instrumental variables in explaining the variation in crimes. As the coefficient of SPI is positive and statistically significant it represents a true causal relationship. I have taken SPI Index as an indicator for income inequality. My expected results are that there should be a positive link between income inequality and crime rates in India. It could be that higher social progress, as measured by SPI, is associated with increased reporting of crimes or changes in law enforcement practices that lead to higher crime detection rates. It's also possible that there are confounding variables or omitted variable bias that are influencing the results. The positive coefficient for population density indicates that an increase in population density is associated with an increase in crimes. Including a lagged value of the dependent variable as an explanatory variable in a regression model can help capture autocorrelation, address endogeneity concerns and here L.crimes is coming positive and significant means previous period's crimes variable influences the current period's crimes variable, after accounting for other explanatory variables in the model

3. Test For Endogeneity

First Approach

By comparing the coefficients from 2SLS and OLS we see that their values are not the same. That implies our model has potential endogeneity issues.

Second Approach

Endogeneity test

```
. estat endog

  Tests of endogeneity
  Ho: variables are exogenous

  Durbin (score) chi2(1)          =  21.6101  (p = 0.0000)
  Wu-Hausman F(1,203)             =   23.286  (p = 0.0000)
```

We reject the null hypothesis here and find that the regressors are endogenous as p value is significant and OLS will give biased estimates. So, we can move ahead and use GMM.

Test For Weak or Strong Instruments

```
. estat firststage

  First-stage regression summary statistics
  --------------------------------------------------------------------------
               |            Adjusted      Partial
      Variable |   R-sq.       R-sq.        R-sq.     F(2,167)   Prob > F
  -------------+------------------------------------------------------------
           spi |  0.4082      0.3833       0.2006      20.9568    0.0000
  --------------------------------------------------------------------------

  Minimum eigenvalue statistic = 20.9568

  Critical Values                      # of endogenous regressors:    1
  Ho: Instruments are weak             # of excluded instruments:     2
  ---------------------------------------------------------------------
                                     |    5%     10%     20%     30%
  2SLS relative bias                 |         (not available)
  -----------------------------------+---------------------------------
                                     |   10%     15%     20%     25%
  2SLS Size of nominal 5% Wald test  |  19.93   11.59    8.75    7.25
  LIML Size of nominal 5% Wald test  |   8.68    5.33    4.42    3.92
  ---------------------------------------------------------------------
```

\

The instruments are weak, shown by low value of R square and the minimum eigenvalue value is also not so high. Suggests that GMM is appropriate as weak instruments are there and I have a greater number of instruments than endogenous regressors.

Overidentification test

```
. estat overid

  Tests of overidentifying restrictions:

  Sargan (score) chi2(1) =  1.00783  (p = 0.3154)
  Basmann chi2(1)        =  .967329  (p = 0.3253)
```

In output, the p-values for both tests are greater than 0.05, which suggests that the null hypothesis of valid instruments cannot be rejected at conventional levels of significance. This indicates that the instruments used in the model are valid, and the 2SLS estimator is unbiased and consistent.

4. Arellano Bover estimation

```
. xtdpdsys crimes pd gr ur ir, lags(1) twostep endog(spi) inst( gsdp edu ) vce(robust) artests(2)

System dynamic panel-data estimation            Number of obs      =       175
Group variable: States                          Number of groups   =        35
Time variable: year
                                                Obs per group:
                                                              min =          5
                                                              avg =          5
                                                              max =          5

Number of instruments =     35                  Wald chi2(6)       =   3304.10
                                                Prob > chi2        =    0.0000
Two-step results
------------------------------------------------------------------------------
             |              WC-Robust
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      crimes |
         L1. |      .2371   .0295977     8.01   0.000     .1790896    .2951105
             |
         spi |   12957.39   2215.873     5.85   0.000     8614.359    17300.42
          pd |   217.4465   25.49573     8.53   0.000     167.4758    267.4172
          gr |   21.94584   297.9857     0.07   0.941    -562.0954    605.9871
          ur |   37.86298   66.28393     0.57   0.568    -92.05113    167.7771
          ir |   1358.366   1505.687     0.90   0.367    -1592.726    4309.458
       _cons |  -714825.9   191656.9    -3.73   0.000     -1090466   -339185.3
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).crimes L(2/.).spi
        Standard: D.pd D.gr D.ur D.ir gsdp edu
Instruments for level equation
        GMM-type: LD.crimes LD.spi
        Standard: _cons
```

L1.crimes, spi and pd have p-value of 0.000, indicating that the lagged crimes variable, spiindex and population density is statistically significant. This suggests that there is a positive relationship between past crimes and current crimes, indicating a potential persistence or feedback effect in crime dynamics, higher social progress index values are associated with higher levels of crimes, suggesting that social progress may not necessarily lead to a decrease in crimes in the given model and an increase in population density is associated with higher levels of crimes.

5. Instrumental variable regression(RE, FE, BE, FD)

Random effects

```
. xtivreg crimes l.crimes pd gr ur ir (spi = edu gsdp), re vce(robust)

G2SLS random-effects IV regression              Number of obs      =       175
Group variable: States                          Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0526                                         min =          5
     between = 0.9521                                         avg =        5.0
     overall = 0.8253                                         max =          5

                                                Wald chi2(6)       =    948.52
corr(u_i, X)       = 0 (assumed)                Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 35 clusters in States)
------------------------------------------------------------------------------
             |               Robust
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         spi |    4614.11   3580.713     1.29   0.198    -2403.959    11632.18
             |
      crimes |
         L1. |   .7084765   .1729291     4.10   0.000     .3695416    1.047411
             |
          pd |    114.513   67.70978     1.69   0.091    -18.19574    247.2217
          gr |  -104.3182   139.0604    -0.75   0.453    -376.8716    168.2351
          ur |  -.8536322    99.1501    -0.01   0.993    -195.1843     193.477
          ir |   3724.059    3304.21     1.13   0.260    -2752.072    10200.19
       _cons |  -169920.2   256590.2    -0.66   0.508    -672827.7    332987.4
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  74991.503
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   spi
Instruments:    L.crimes pd gr ur ir edu gsdp
------------------------------------------------------------------------------
```

Fixed effects

```
. xtivreg crimes l.crimes pd gr ur ir (spi = edu gsdp), fe vce(robust)

Fixed-effects (within) IV regression            Number of obs      =       175
Group variable: States                          Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0951                                         min =          5
     between = 0.6830                                         avg =        5.0
     overall = 0.6213                                         max =          5


                                                Wald chi2(6)       =    390.26
corr(u_i, Xb)  = -0.0789                        Prob > chi2        =    0.0000

                                 (Std. Err. adjusted for 35 clusters in States)
------------------------------------------------------------------------------
             |               Robust
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         spi |   10520.86   9400.822     1.12   0.263    -7904.411    28946.13
             |
      crimes |
         L1. |   .0675921   .1300961     0.52   0.603    -.1873916    .3225758
             |
          pd |   310.6049   317.3877     0.98   0.328    -311.4636    932.6734
          gr |    396.409   336.1548     1.18   0.238    -262.4423     1055.26
          ur |   224.1305   150.4041     1.49   0.136    -70.65617    518.9172
          ir |   3485.875   2231.371     1.56   0.118    -887.5322    7859.282
       _cons |  -970498.2   659433.1    -1.47   0.141     -2262963    321966.9
-------------+----------------------------------------------------------------
     sigma_u |   115168.7
     sigma_e |  74991.503
         rho |   .70225232   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   spi
Instruments:    L.crimes pd gr ur ir edu gsdp
------------------------------------------------------------------------------
```

Between effects

```
. xtivreg crimes l.crimes pd gr ur ir (spi = edu gsdp), be vce(robust)

Between-effects IV regression:                  Number of obs      =       175
Group variable: States                          Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0319                                         min =         5
     between = 0.9941                                         avg =       5.0
     overall = 0.8257                                         max =         5


                                                Wald chi2(6)       =   3562.33
sd(u_i + avg(e_i.))=  17198.65                  Prob > chi2        =    0.0000

                                  (Std. Err. adjusted for 35 clusters in States)

------------------------------------------------------------------------------
             |               Robust
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         spi |   894.9336   1020.676     0.88   0.381    -1105.555    2895.422
             |
      crimes |
         L1. |   1.014913   .0795176    12.76   0.000     .8590614    1.170764
             |
          pd |    10.9998   25.82114     0.43   0.670    -39.60871    61.60831
          gr |  -126.1017   108.3835    -1.16   0.245    -338.5294      86.326
          ur |  -31.98759     60.526    -0.53   0.597    -150.6164    86.64119
          ir |  -269.1271   3182.158    -0.08   0.933    -6506.042    5967.788
       _cons |   74023.88   78095.57     0.95   0.343    -79040.62    227088.4
------------------------------------------------------------------------------
Instrumented:   spi
Instruments:    L.crimes pd gr ur ir edu gsdp
------------------------------------------------------------------------------
```

First Difference

```
. xtivreg crimes l.crimes pd gr ur ir (spi = edu gsdp), fd vce(robust)

First-differenced IV regression
Group variable:     States                      Number of obs      =       140
Time variable:      year                        Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0002                                         min =         4
     between = 0.0225                                         avg =       4.0
     overall = 0.2526                                         max =         4


                                                Wald chi2(6)       =    161.96
corr(u_i, Xb)   = -0.6540                       Prob > chi2        =    0.0000

                                  (Std. Err. adjusted for 35 clusters in States)

------------------------------------------------------------------------------
             |               Robust
    D.crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         spi |
         D1. |   5134.521   18253.43     0.28   0.778    -30641.55    40910.59
             |
      crimes |
         LD. |  -.6340431   .1508183    -4.20   0.000    -.9296415   -.3384446
             |
          pd |
         D1. |    197.793   185.1797     1.07   0.285    -165.1525    560.7385
             |
          gr |
         D1. |   216.0749   234.7729     0.92   0.357    -244.0716    676.2213
             |
          ur |
         D1. |   197.3972   114.8309     1.72   0.086    -27.66723    422.4615
             |
       _cons |   9368.218   16211.64     0.58   0.563    -22406.02    41142.45
-------------+----------------------------------------------------------------
     sigma_u |  262564.51
     sigma_e |  170718.05
         rho |  .70286263   (fraction of variance due to u_i)
------------------------------------------------------------------------------
Instrumented:   spi
Instruments:    L.crimes pd gr ur ir edu gsdp
------------------------------------------------------------------------------
```

From the IV regression from different categories, it can be seen that First difference approach is the most appropriate one. Here D(l.crimes) comes significant and also D.ur comes significant at 10% significance level. The difference of lagged crime rate variable has a statistically significant relationship with the outcome variable D.crime. The negative sign of the coefficient (-0.6340431) suggests that an increase in the lagged crime rate is associated with a decrease in the outcome variable (assuming all other variables are held constant). This suggests that there is some evidence of a relationship between the first-differenced unemployment rate variable (D1(ur)) and the outcome variable of interest, as estimated using the instrumental variable (IV) regression with the given instruments, but the evidence is not as strong compared to the more conventional significance levels.The positive sign of the coefficient (197.3972) suggests that an increase in the first-differenced unemployment rate is associated with an increase in the outcome variable.

6. **Threshold Regression**

```
. xthenreg crimes pd gr ur ir, endogenous( spi) inst( edu gsdp)

N = 35, T = 6
Panel Var. = States
Time Var. = year
Number of moment conditions = 44
```

crimes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Lag_y_b	-.0496947	.0207857	-2.39	0.017	-.090434	-.0089554
gr_b	-150.841	94.43587	-1.60	0.110	-335.9319	34.24992
ur_b	-77.34815	25.31566	-3.06	0.002	-126.9659	-27.73037
ir_b	-1582.916	343.8279	-4.60	0.000	-2256.806	-909.0253
spi_b	477.6388	970.3998	0.49	0.623	-1424.31	2379.587
cons_d	-4036691	301437.3	-13.39	0.000	-4627497	-3445885
Lag_y_d	-.3807412	.0345938	-11.01	0.000	-.4485438	-.3129386
gr_d	3112.565	256.8075	12.12	0.000	2609.231	3615.898
ur_d	602.3723	71.33512	8.44	0.000	462.558	742.1865
ir_d	7781.794	1941.989	4.01	0.000	3975.565	11588.02
spi_d	24325.9	1805.221	13.48	0.000	20787.73	27864.07
r	676.153	67.07342	10.08	0.000	544.6916	807.6145

For estimating dynamic effects of variables we use xthenreg command in stata, that is Dynamic panel-data model allowing threshold and endogeneity (regression). Here _b implies how the variable will effect crimes at low regime level(threshold level) and _d implies how variables will effect crimes at high regime levels. r is the threshold rate here. Spi, gr, ur, ir and lag of crimes has a significant effect above threshold level. But below threshold level only lag crime rates, uer and ir are significant. The significant threshold level is 676.153 here

7. Unit root tests

For LLC (Levin-Lin-Chu (LLC) test) if p value is significant we reject the null hypothesis of unit roots and confirm that the root is stationary.

Fisher-type unit-root test uses the augmented Dickey-Fuller (ADF) test and its null hypothesis (Ho) in this case is that all panels contain unit roots, and the alternative hypothesis (Ha) is that at least one panel is stationary,

```
. xtunitroot llc crimes

Levin-Lin-Chu unit-root test for crimes
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t      -20.6578
 Adjusted t*       -21.8716       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc spi

Levin-Lin-Chu unit-root test for spi
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t      -1.1e+03
 Adjusted t*       -1.2e+03       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc edu

Levin-Lin-Chu unit-root test for edu
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t      -33.3811
 Adjusted t*       -36.0451       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc gr

Levin-Lin-Chu unit-root test for gr
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t       -7.6007
 Adjusted t*        -6.6126       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc gsdp

Levin-Lin-Chu unit-root test for gsdp
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t      -15.0261
 Adjusted t*       -14.6201       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc ur

Levin-Lin-Chu unit-root test for ur
---------------------------------------
Ho: Panels contain unit roots            Number of panels  =     35
Ha: Panels are stationary                Number of periods =      6

AR parameter: Common                     Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                  Statistic      p-value
------------------------------------------------------------------
 Unadjusted t       -7.8920
 Adjusted t*        -6.7405       0.0000
------------------------------------------------------------------
```

```
. xtunitroot llc ir

Levin-Lin-Chu unit-root test for ir
------------------------------------
Ho: Panels contain unit roots              Number of panels  =     35
Ha: Panels are stationary                  Number of periods =      6

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------
 Unadjusted t        -37.8491
 Adjusted t*         -40.5733       0.0000
------------------------------------------------------------------
```

```
. xtunitroot fisher l.crimes, dfuller lags(0)
(35 missing values generated)

Fisher-type unit-root test for L.crimes
Based on augmented Dickey-Fuller tests
---------------------------------------
Ho: All panels contain unit roots           Number of panels  =     35
Ha: At least one panel is stationary        Number of periods =      5

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included
Drift term:   Not included                  ADF regressions: 0 lags
--------------------------------------------------------------------------
                                    Statistic      p-value
--------------------------------------------------------------------------
 Inverse chi-squared(70)    P        132.6315       0.0000
 Inverse normal             Z         -0.4489       0.3268
 Inverse logit t(179)       L*        -1.6290       0.0525
 Modified inv. chi-squared  Pm         5.2933       0.0000
--------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
--------------------------------------------------------------------------
```

Here all the above roots- crimes,l.crimes spi, edu, gsdp, gr, ur and ir are stationary as their p value is highly significant

```
. xtunitroot llc pd

Levin-Lin-Chu unit-root test for pd
------------------------------------
Ho: Panels contain unit roots              Number of panels  =     35
Ha: Panels are stationary                  Number of periods =      6

AR parameter: Common                       Asymptotics: N/T -> 0
Panel means:  Included
Time trend:   Not included

ADF regressions: 1 lag
LR variance:     Bartlett kernel, 5.00 lags average (chosen by LLC)
------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------
 Unadjusted t          0.1073
 Adjusted t*           0.1398       0.5556
------------------------------------------------------------------
```

```
. xtunitroot fisher dpd, dfuller lags(0)
(35 missing values generated)

Fisher-type unit-root test for dpd
Based on augmented Dickey-Fuller tests
---------------------------------------
Ho: All panels contain unit roots           Number of panels  =     35
Ha: At least one panel is stationary        Number of periods =      5

AR parameter: Panel-specific                Asymptotics: T -> Infinity
Panel means:  Included
Time trend:   Not included
Drift term:   Not included                  ADF regressions: 0 lags
--------------------------------------------------------------------------
                                    Statistic      p-value
--------------------------------------------------------------------------
 Inverse chi-squared(70)    P        259.9818       0.0000
 Inverse normal             Z         -7.3269       0.0000
 Inverse logit t(179)       L*       -11.1774       0.0000
 Modified inv. chi-squared  Pm        16.0564       0.0000
--------------------------------------------------------------------------
 P statistic requires number of panels to be finite.
 Other statistics are suitable for finite or infinite number of panels.
--------------------------------------------------------------------------
```

pd variable that is population density is nonstationary. But taking its first difference(dpd) changes it into stationary root.

8. ARDL Method

BOUNDS TEST

```
> X<-cbind(SPI, PD, EDU, GR, GSDP, UR, IR )
> pdata<-pdata.frame(ECO342_200307_DATASET, index= c("STATES", "YEAR"))
> formula<-CRIMES~SPI+PD+EDU+GR+GSDP+UR+IR+lag(CRIMES,1)+lag(SPI,1)
> bounds_test<-pbgtest(formula, data=pdata, order=1)
> summary(bounds_test)
             Length Class  Mode
statistic      1    -none- numeric
parameter      1    -none- numeric
method         1    -none- character
p.value        1    -none- numeric
data.name      1    -none- character
coefficients  11    -none- numeric
vcov         121    -none- numeric
alternative    1    -none- character
> bounds_test$p.value
[1] 1.401256e-06
```

p-value from the bounds test is significant, it suggests evidence in favor of the presence of a long-run relationship between the variables and suggests that cointegration may be present in the data. Evidence of cointegration between variables suggests applying cointegration analysis techniques, such as error correction models (ECMs, to estimate the long-run relationship and short-run dynamics among the variables.

ARDL TEST IN STATA

First check which all variables are significant by running panel regression

pd and gsdp are significant here

```
. xtreg crimes L.crimes spi pd edu gr gsdp ur ir

Random-effects GLS regression                   Number of obs      =       175
Group variable: states                          Number of groups   =        35

R-sq:                                           Obs per group:
     within  = 0.0493                                          min =         5
     between = 0.9799                                          avg =       5.0
     overall = 0.8427                                          max =         5

                                                Wald chi2(8)       =    889.09
corr(u_i, X)   = 0 (assumed)                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
      crimes |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      crimes |
         L1. |   .7521923    .055103    13.65   0.000     .6441924    .8601922
             |
         spi |     528.65   970.9643     0.54   0.586    -1374.405    2431.705
          pd |   60.29692   26.18302     2.30   0.021     8.979139    111.6147
         edu |  -1042.287   758.1131    -1.37   0.169    -2528.161    443.5875
          gr |   34.40321   136.5195     0.25   0.801    -233.1701    301.9765
        gsdp |   .0003497   .0001854     1.89   0.059    -.0000138    .0007131
          ur |  -117.4186   169.5974    -0.69   0.489    -449.8234    214.9862
          ir |   4076.941    3351.42     1.22   0.224    -2491.721     10645.6
       _cons |   21089.51   132244.4     0.16   0.873    -238104.7    280283.8
-------------+----------------------------------------------------------------
     sigma_u |          0
     sigma_e |  74536.263
         rho |          0   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

xtpmg- the ardl cointegration approach using mg(mean group error correction model) and dfe(dynamic fixed effect regression). In stata, there are two main commands used for estimating ardl models: mg (for the general ardl model) and dfe (for the dynamic fixed-effects ardl model).

MG

```
. xtpmg d.crimes d.spi d.pd d.gsdp, lr(l.crimes spi pd gsdp) ec(ECT) replace mg
```

Mean Group Estimation: **Error Correction Form**
(Estimate results saved as **mg**)

D.crimes	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ECT						
spi	25.30538	25.30538	1.00	0.317	-24.29226	74.90302
pd	1585.149	1165.368	1.36	0.174	-698.9302	3869.229
gsdp	-.0002931	.0012265	-0.24	0.811	-.0026971	.0021108
SR						
ECT	-1.002738	.3902149	-2.57	0.010	-1.767545	-.237931
spi						
D1.	11.72753	12.77673	0.92	0.359	-13.31439	36.76946
pd						
D1.	1584.628	1158.619	1.37	0.171	-686.2227	3855.479
gsdp						
D1.	-.0112804	.0113008	-1.00	0.318	-.0334295	.0108687
_cons	320037.6	466603.4	0.69	0.493	-594488.4	1234563

Results interpretation-the estimated coefficients under the "ECT" section represent the long-run relationship between the variables. In this case, the coefficients for the variables spi, pd and gsdp represent the long-run effects of changes in these variables on the dependent variable D.crimes. However, based on the p-values (i.e., the "P>|z|" column), none of these coefficients are statistically significant at conventional significance levels (e.g., α = 0.05), as all of them have p-values above 0.05. This means that there is not enough evidence to support the presence of a statistically significant long-run relationship between spi, pd or gsdp and D.crimes.

The "SR" section represents the short-run dynamics, where ECT refers to the error correction term, and "D1." represents the first difference of the variables. The coefficient for the error correction term ("ECT") is statistically significant at a significant level of 0.05, as its p-value is below 0.05 (i.e., 0.010). This suggests that there is a statistically significant short-run relationship between the error correction term and D.crimes, indicating that the model is capturing short-run dynamics in the data. The coefficients for the first differences of the variables (e.g., "D1.spi", "D1.pd", and "D1.gsdp") are not statistically significant at conventional significance levels, as their p-values are above 0.05. This suggests that there is no statistically significant short-run relationship between the first differences of these variables and "D.crimes".

DFE

```
. xtpmg d.crimes d.spi d.pd  d.gsdp, lr(l.crimes spi pd  gsdp) ec(ECT) replace dfe

Dynamic Fixed Effects Regression: Estimated Error Correction Form
(Estimate results saved as DFE)
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ECT						
spi	-2903.019	5387.874	-0.54	0.590	-13463.06	7657.021
pd	457.8696	450.7589	1.02	0.310	-425.6016	1341.341
gsdp	.0025508	.0011244	2.27	0.023	.000347	.0047546
SR						
ECT	-.9189008	.0855085	-10.75	0.000	-1.086494	-.7513073
spi D1.	-950.4587	5586.429	-0.17	0.865	-11899.66	9998.74
pd D1.	-401.9203	398.5402	-1.01	0.313	-1183.045	379.204
gsdp D1.	-.0047256	.0014864	-3.18	0.001	-.0076388	-.0018124
_cons	18176.03	256727.9	0.07	0.944	-485001.4	521353.5

Based on the results provided, the variable gsdp have a statistically significant long-run effect on "D.crimes. Changes in spi and pd are not reliably associated with changes in D.crimes in the long-run as they are not statistically significant. In terms of short-run effects, the error correction term (ECT) variable is statistically significant and suggests that there is a statistically significant short-run relationship between the error correction term and "D.crimes", indicating that the model is capturing short-run dynamics in the data. Simalarily D.gsdp has a short run relationship with D.crimes

hausman mg DFE

```
. hausman mg DFE

Note: the rank of the differenced variance matrix (2) does not equal the number of coefficients being tested (3); be sure
      this is what you expect, or there may be problems computing the test.  Examine the output of your estimators for
      anything unexpected and possibly consider scaling your variables so that the coefficients are on a similar scale.
```

	Coefficients (b) mg	(B) DFE	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
spi	25.30538	-2903.019	2928.324	.
pd	1585.149	457.8696	1127.279	1074.662
gsdp	-.0002931	.0025508	-.0028439	.00049

```
                    b = consistent under Ho and Ha; obtained from xtpmg
       B = inconsistent under Ha, efficient under Ho; obtained from xtpmg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =        0.86
                Prob>chi2 =      0.6514
                (V_b-V_B is not positive definite)
```

Based on the results of the Hausman test, the chi2 is positive and the p-value (Prob>chi2) is 0.6514, which is greater than the conventional significance level of 0.05. Therefore, we fail to reject the null hypothesis (Ho) that the difference in coefficients between the "mg" and "DFE" estimators is not systematic. One can go with mg method here which is consistent under Ho.

Objective And Significance

This study's goal is to use panel data analysis to look into the connection between income inequality and crime in India. The study specifically seeks to determine whether income inequality has an impact on the prevalence of crime in India and to identify the ways in which it does. Additionally, the study will examine how socioeconomic and demographic factors for ex-age, gender, education level, and unemployment influence the link between income inequality and crime. Using proper econometric methods such as GMM estimation and fixed difference regression, the study will also address the issue of endogeneity. Also, the study will use regression models to evaluate the non-linear association between income inequality and crime using threshold regression. This study will advance the field of literature by providing new insights into the relationship between income inequality and property and violent crime in India. Moreover, the study's findings could have significant policy implications for crime prevention and reduction in India. Policymakers could think about introducing policies to minimize economic disparity, such as progressive taxation, social welfare programs, and labor market changes if it is discovered that income inequality significantly predicts crime. The decrease in crime and the improvement of social welfare in India may both benefit from these approaches. Additionally, understanding the non-linear relationship between income inequality and crime could inform the development of targeted interventions to reduce crime among vulnerable groups in society.

Conclusions

We analyzed the relationship between demographic, education, unemployment, income inequality, and crime rates using the Indian panel dataset, and gained insights into the specific factors that contribute to crime in India. It showed a positive link between income inequality and crime rates in the nation. And population density and GDSP has a positive effect on crime rate .By addressing the endogeneity issue, the research provided robust evidence of the causal effect of income inequality on crime. The use of a variety of econometric techniques, such as the panel VAR model, threshold, GMM, ARDL Cointegration, helped to identify the precise nature of the relationship between income inequality and crime rates in India. The research provided essential insights into the mechanisms through which income inequality affects crime rates in the country and can contribute to the development of policies aimed at reducing income inequality and promoting social welfare.

Data Sources

The data for SPI Index was collected from Press Information Bureau, collected by EAC-PM for all Indian states and UTs. Crime rates were collected from NCRB site. Gender ratio was gathered from Census data. Net enrollment rate for primary education were collected from UDISE+ reports by Department of school Education and Literacy, Ministry of education, Government of India. RBI site, Handbook of Statistics of Indian states provided data on population density and GSDP. Ministry of Statistics and Programme Implementation (MoSPI) and National Sample Survey Office (NSSO) provided data on inflation rate. Centre for Monitoring Indian Economy (CMIE), NSSO and Ministry of Labour and Employment, Government of India were used to gather unemployment rate. Various web search results also helped compile the dataset.

References

Paul-Philippe Pare, and Richard Felson. "Income Inequality, Poverty, and Crime Across Nations." The British Journal of Sociology 2014 Volume 65 Issue 3.Sept 2014.

Joseph Deutsch, Uriel Spiegel & Joseph Templeman.. Crime and income inequality:An economic approach". Atlantic Economic Journal. December 1992.

Morgan Kelly."Inequality and Crime". Review of Economics and Statistics. February 2000.

LIM, THYE GOH, SHIVEE RANJANEE KALIAPPAN, AND SURYATI ISHAK. "Income Inequality and Crime: Evidence from a Dynamic Panel Data". International Journal of Economics and Management. 2018

Brush, Jesse. "Does Income Inequality Lead to More Crime? A Comparison of Cross-Sectional and Time-Series Analyses of United States Counties.", 2007

Ted Enamorado , Luis F. López-Calva, Carlos Rodríguez-Castelán, and Hernán Winkler. "Income Inequality and Violent Crime: Evidence from Mexico's Drug War.", Journal of Development Economics, Volume 120, May 2016