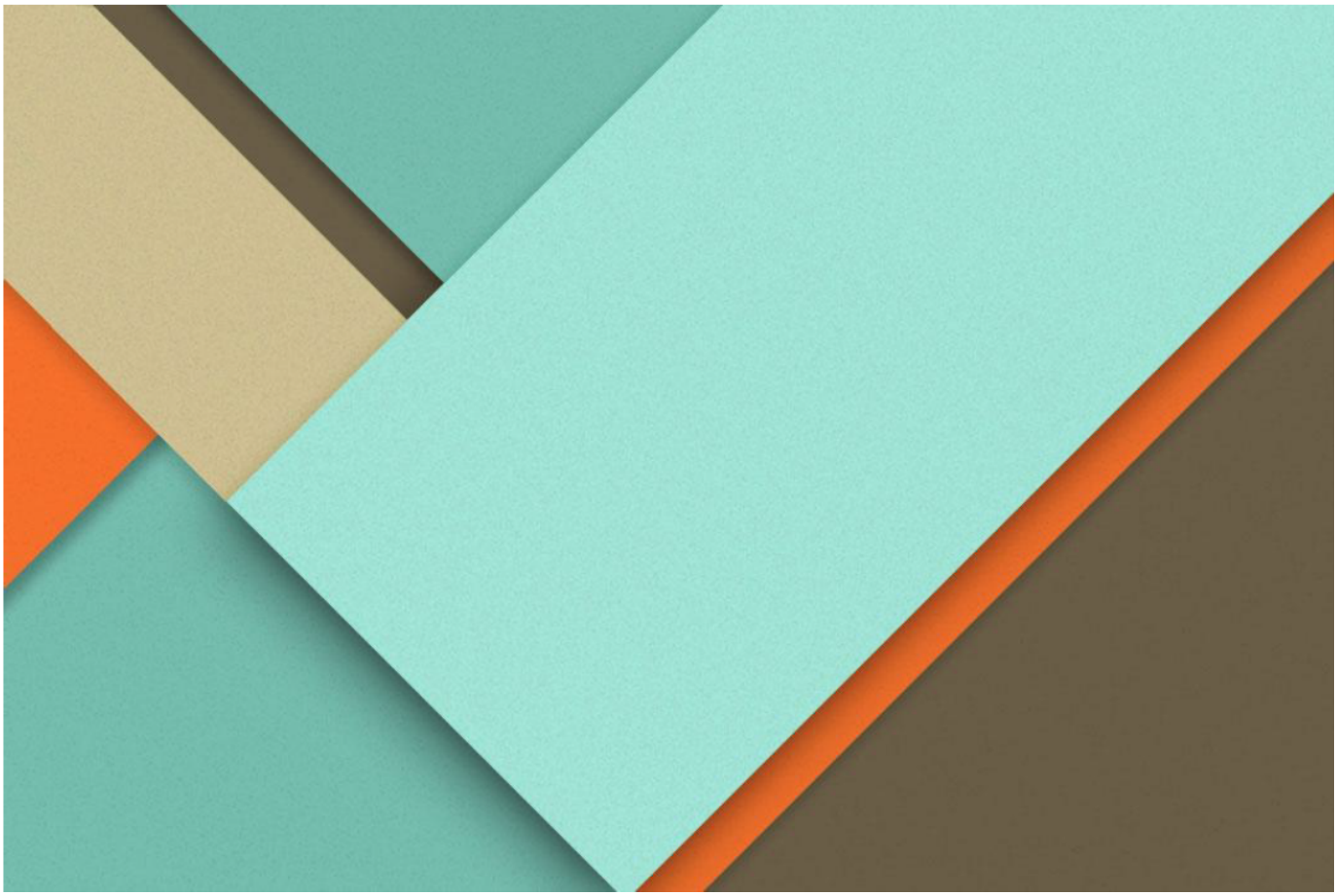

Analysis of Poverty in India

OBJECTIVE

In this report, we aim to provide an analysis of the determinants of the poverty rate and how various factors influence it across all the states of India (for the year 2011). Poverty is a major issue in India and additional measures are needed to tackle this phenomenon. We expect that after observing the results, we might have a better understanding of the appropriate measures required to reduce poverty, mainly by seeing its causes and then applying necessary solutions to reduce it. We expect our results to be in accordance with the theoretical literature.

SPECIFICATION

We chose the following independent variables we thought would affect the Poverty Rate --

- Total Population
- Population Density
- Death Rate
- Literacy Rate
- Birth Rate
- Infant Mortality Rate
- Unemployment Rate
    - Rural
    - Urban
- Expenditure on Social Sector
- Inflation Rate
- Monthly Per Capita Expenditure
    - Rural
    - Urban
- Gross State Domestic Product
- GDP (State) growth from previous year

Our Dependent Variable (Poverty Rate) was of the year 2011.

So we took data of some independent for the year 2011 whereas we took data of some independent variables for earlier years (2009, 2010) as well to accommodate for the lag effect these factors may have on our dependent variable

For Independent Variable we took -

Total Population, Literacy Rate, Inflation Rate, Population Density - 2011

Death Rate, Birth Rate, Infant Mortality Rate - 2009, 2010, 2011

Expenditure on Social Sector - 2009/10, 2010/11

Unemployment Rate - 2009/10

Monthly Per Capita Expenditure, Gross State Domestic Product, GDP (State) growth from previous year - 2011

Hence our final model is:

$\text{Poverty Rate}_{2011} = f(\text{Inflation Rate}_{2011}, \text{Birth\_Rate}_{2010}, \text{Rural\_Unemployment}_{2009\text{-}10}, \text{Average\_MPCE}_{2009\text{-}10})$

MEASUREMENT OF VARIABLES

I. Dependent Variable

Poverty_Rate_2011 - Percentage of people living below the poverty line for the year 2011. The poverty line is taken as half the median household income of the total population.

II. Independent Variable

Total_Population_2011 - The total population in a particular region for the year 2011.

Death_rate_t - The number of deaths per 1000 people in a particular region per year for the year t (t=2009, 2010, 2011).

Literacy_rate_2011- Percentage of the population above the age of 7 that can read and write for the year 2011.

Birth_rate_t - The number of births per 1000 people in a particular region per year for the year t (t=2009, 2010, 2011).

Infant_mortality_rate_t - The number of infant deaths below the age of one per 1000 live births in a particular region per year for the year t (t=2009, 2010, 2011)

Urban_unemployment_rate_2009-10 - The number of people who are above the age of 15, living in urban areas and unemployed per 1000 people for the period 2009-10.

Rural_unemployment_rate_2009-10 - The number of people who are above the age of 15, living in rural areas and unemployed per 1000 people for the period 2009-10.

Social_spending_t - The expenditure of the government on the social sector which includes housing, rural and urban development, the welfare of SC, STs, and OBCs, etc in Rs Crore for the period t (t=2009-10, 2010-11).

Inflation_rate_2011 - An overall increase in the Consumer Price Index (CPI), which is a weighted average of prices for different goods, in one year for the year 2011.

Monthly_per_capita_expenditure_urban_2011 - Refers to the monthly expenditure in Rs. of a person living in an urban area for the year 2011.

Monthly_per_capita_expenditure_rural_2011 - Refers to the monthly expenditure in Rs. of a person living in a rural area for the year 2011.

Gross_state_Product_at_constant_prices_2011-12 - The sum total volume of all finished goods and services produced during the year 2011-12 within the geographical boundaries of the State, accounted for without duplication at market prices using the base year 2004-05.

Growth_from_Previous_Year - Percentage growth of GSDP at constant prices from the previous year.

Population_density_2011 - Number of people living in a region per square km.

SOURCES OF DATA

- RBI (Handbook of Statistics)
- Ministry of Statistics and Program Implementation
- Economic Survey of India

IMPACT OF INDEPENDENT VARIABLES

We expected the following impact of our independent variables

Total Population - We expect a negative impact, as more population would mean fewer resources for everyone

Population Density - We expect a negative impact, as more population density would mean fewer resources for everyone

Death Rate - We expect a positive impact

Literacy Rate - We expect a negative impact, as a higher literacy rate could mean better education and thus more earning opportunities

Birth Rate - We expect a positive impact, as a higher birth rate would mean more division of resources in family

Infant Mortality Rate - We expect a positive impact, since it means poor health infrastructure

Unemployment Rate - We expect a positive impact, as more unemployment would mean less income and thus more poverty, both in Rural & Urban

Expenditure on Social Sector - We expect a negative impact, as more expenditure by the government on the social sector would help improve the quality of life for people

Inflation Rate - We expect a positive impact, as more inflation would mean even basic necessities of life could become more costly

Monthly Per Capita Expenditure - We expect a negative impact, as an individual tends to spend more on consumption for higher income.

- Rural
- Urban

Gross State Domestic Product - We expect a negative impact, as higher production in the state would mean more income opportunities and thus less poverty.

GDP (State) growth from previous year - We expect a negative impact, as growth in production in the state would mean more income opportunities and thus less poverty

ESTIMATION TECHNIQUE & PROCESS FOLLOWED

We estimate our model using the method of OLS. We used the following independent variables in our initial baseline model: Total_Population_2011, Death_Rate_2011, Infant_Mortality_Rate_2011, Birth_Rate_2011, Social_Spending_2011_12, Inflation_rate_2011. The estimated model gave a very poor fit with the coefficients of Death_Rate_2011, Infant_Mortality_Rate_2011, Social_Spending_2011_12 being statistically insignificant. We suspected that perhaps the infant mortality rate, death rate, and birth rate could have a lagged effect on the poverty rate. So we took the infant mortality rate, death rate, and birth rate for the year 2010 in our next model. Birth rate for the year 2010 still comes out to be statistically significant with a higher coefficient than that for the year 2011. But infant mortality rate and death rate still prove to be statistically insignificant. This is also the case while taking the same variables for the year 2009. We also took social spending for the year 2009-10 as it may take time for the government's spending to have an impact on poverty. But it still comes out to be statistically insignificant.

We also tested for inclusion of variables Literacy_rate_2011, rural_unemployment_2011/12, urban_unemployment_2011/12, urban_unemployment_2009/10, rural_unemployment_2009/10, MPCE_for_rural_2011/12, MPCE_for_urban_2011/12, Gross_Domestic_Product_2011/12_at_constant_prices_2004/05, growth from previous year and Population_density_2011 using Restricted F-test separately for each variables. We found the coefficients of rural_unemployment_2009/10, MPCE_for_rural_2011/12, MPCE_for_urban_2011/12, to be statistically significant. Surprisingly the coefficient of

Literacy_rate_2011 came out to be statistically insignificant. We see that the coefficient of rural_unemployment_2009/10 is statistically significant but that of 2011/12 is not indicating lagged effect on the dependent variable. However this does not hold for urban_unemployment.

Hence our statistically significant model included Inflation_rate_2011, Birth_rate_2010, rural_unemployment_2009/10, MPCE_for_rural_2011/12, MPCE_for_urban_2011/12 as the independent variables.

```
. reg povertyrate_2011 birthrate2010 inflation_rate_201112 mpceforrural201112 mpceforurban201112 ruralunemployment200910

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  5,    30) =   26.21
       Model |  3718.83236     5  743.766473            Prob > F      =  0.0000
    Residual |  851.255257    30  28.3751752            R-squared     =  0.8137
-------------+------------------------------           Adj R-squared =  0.7827
       Total |  4570.08762    35  130.573932            Root MSE      =  5.3268

------------------------------------------------------------------------------
povertyrate_2011 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
     birthrate2010 |   .5910319   .2761422     2.14   0.041     .0270744    1.154989
inflation_rate_201112 |  -1.265737   .2796945    -4.53   0.000    -1.836949   -.6945242
mpceforrural201112 |  -.0110074   .0029351    -3.75   0.001    -.0170017    -.005013
mpceforurban201112 |  -.0057945   .0022926    -2.53   0.017    -.0104767   -.0011124
ruralunemployment200910 |   .0959786   .0237244     4.05   0.000     .0475269    .1444303
           _cons |   48.23866   9.593095     5.03   0.000     28.64695    67.83037
------------------------------------------------------------------------------
```

We'll now perform various tests such as tests for multicollinearity, heteroscedasticity, and testing for normality of random disturbance term. Also since the data is available only over a very few years, we need not test for autocorrelation or structural breaks.

Jarque - Bera Test for Normality:

We conducted this test to check if the random disturbance term follows normal distribution. After observing the result we failed to reject the null hypothesis and thus the random disturbance term follows normal distribution.

```
. jb6 resid
Jarque-Bera normality test:   1.03 Chi(2)   .5974
Jarque-Bera test for Ho: normality: (resid)
```

Restricted F test:

We'll also test for equality of coefficients using Restricted F-test to check whether two particular coefficients are statistically significantly different. We performed such tests for Birth_rate_2010 and rural unemployment_2009/10; and average_mpce and inflation_rate_2011. On performing F-test our null hypothesis is rejected for both the tests indicating that the coefficients of all four of them are statistically significantly different

```
. test birthrate2010 = ruralunemployment200910

 ( 1)  birthrate2010 - ruralunemployment200910 = 0

       F(  1,    31) =    4.42
            Prob > F =    0.0438

. test inflation_rate_201112 = avg_mpce

 ( 1)  inflation_rate_201112 - avg_mpce = 0

       F(  1,    31) =   19.82
            Prob > F =    0.0001
```

Multicollinearity Test:

Multicollinearity is caused due to reasons such as inclusion of variables computed from other variables in the equation, inclusion of similar variables more than once etc. It is important to reduce the extent of multicollinearity as it has various implications like R square can be very high and greater multicollinearity raises standard errors.

In our model after computing the regression results and individual significance statistical tests, we move towards checking multicollinearity. A low value of VIF indicated that multicollinearity is not severe.

```
. vif
```

Variable	VIF	1/VIF
mpceforrur~2	3.11	0.321123
mpceforurb~2	2.22	0.450237
birthra~2010	1.81	0.551198
rural~200910	1.43	0.700498
infla~201112	1.10	0.910229
Mean VIF	1.94	

To confirm this, we performed further analysis using the Farrar Glaubar test. The results from the Farrar Glaubar test showed that multicollinearity existed and in order to reduce it we performed linear transformation on the variables MPCE rural and MPCE urban. A new variable was generated average MPCE which was the average MPCE of both urban and rural for that particular state.

```
. fgtest povertyrate_2011 birthrate2010 inflation_rate_201112 mpceforrural201112 mpceforurban201112 ruralunemployment200910

=============================================================================
* Farrar-Glauber Multicollinearity Tests
=============================================================================
  Ho: No Multicollinearity - Ha: Multicollinearity

* (1) Farrar-Glauber Multicollinearity Chi2-Test:
    Chi2 Test =    55.1882     P-Value > Chi2(10) 0.0000

* (2) Farrar-Glauber Multicollinearity F-Test:
```

Variable	F_Test	DF1	DF2	P_Value
birth~2010	6.310	31.000	4.000	0.042
inf~201112	0.764	31.000	4.000	0.711
mpceforr~2	16.384	31.000	4.000	0.007
mpceforu~2	9.463	31.000	4.000	0.020
rur~200910	3.314	31.000	4.000	0.125

```
* (3) Farrar-Glauber Multicollinearity t-Test:
```

Variable	b~2010	infl~2	mpce..	mpce..	rura~0
bir~2010	.				
i~201112	-0.887	.			
m~rura~2	-4.754	0.374	.		
m~urba~2	-3.226	-0.090	5.902	.	
r~200910	-2.480	1.539	2.869	1.271	.

After the linear transformation, we found a significant decrease in the VIF value and thus we have successfully decreased the severity of multicollinearity.

```
. reg povertyrate_2011 birthrate2010 inflation_rate_201112 avg_mpce ruralunemployment200910

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  4,    31) =   32.23
       Model |  3684.21053     4  921.052633            Prob > F      =  0.0000
    Residual |  885.877088    31  28.5766803            R-squared     =  0.8062
-------------+------------------------------           Adj R-squared =  0.7811
       Total |  4570.08762    35  130.573932            Root MSE      =  5.3457

------------------------------------------------------------------------------
povertyrate_2011 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
     birthrate2010 |   .6510034    .271712     2.40   0.023     .0968431    1.205164
inflation_rate_201112 |  -1.264356   .2806831    -4.50   0.000    -1.836813   -.6918994
           avg_mpce |  -.0160157   .0022361    -7.16   0.000    -.0205763   -.0114552
ruralunemployment200910 |   .0882685   .0227547     3.88   0.001     .0418599    .1346771
               _cons |   48.00362   9.624729     4.99   0.000     28.37386    67.63339
------------------------------------------------------------------------------
```

```
. vif

    Variable |       VIF       1/VIF
-------------+----------------------
birthra~2010 |      1.74    0.573362
    avg_mpce |      1.68    0.595633
rural~200910 |      1.30    0.766880
infla~201112 |      1.10    0.910247
-------------+----------------------
    Mean VIF |      1.46
```

Heteroscedasticity Test:

Heteroscedasticity arises when the variance of the random disturbance term is not constant across observations. It usually arises in cross-sectional data, but may also be present in time-series data. Since our given data is a mixture of cross-sectional and time-series data, there is a possibility that the problem of heteroscedasticity may be present.

We perform the Breusch-Pagan test to check for heteroscedasticity which has constant variance across observations as its null hypothesis. On performing the test we get a very high p-value and hence fail to reject our null hypothesis. Hence our estimated model does not suffer from the problem of heteroscedasticity.

To reaffirm our results, we also perform White's test, which the null hypothesis that our estimated model has homoscedasticity. We again get a p-value greater than 0.1 and hence fail to reject the null-hypothesis.

Hence we can conclude that the estimated model does not suffer from the problem of heteroscedasticity.

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of povertyrate_2011

         chi2(1)      =     0.06
         Prob > chi2  =   0.8001

. estat hettest birthrate2010 inflation_rate_201112 avg_mpce ruralunemployment200910

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: birthrate2010 inflation_rate_201112 avg_mpce ruralunemployment200910

         chi2(4)      =     1.17
         Prob > chi2  =   0.8831

. estat imtest, white

White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(14)     =     8.62
         Prob > chi2  =   0.8547

Cameron & Trivedi's decomposition of IM-test

              Source |       chi2     df       p
---------------------+---------------------------
  Heteroskedasticity |       8.62     14    0.8547
            Skewness |       2.42      4    0.6589
            Kurtosis |       0.87      1    0.3510
---------------------+---------------------------
               Total |      11.91     19    0.8895
```

.

RESULT & IMPLICATION

Our final model comes out to be -

Poverty Rate$_{2011}$ = f(Inflation Rate$_{2011}$, Birth_Rate$_{2010}$, Rural_Unemployment$_{2009\text{-}10}$, Average_MPCE$_{2009\text{-}10}$)

```
. reg povertyrate_2011 birthrate2010 inflation_rate_201112 avg_mpce ruralunemployment200910

      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  4,    31) =   32.23
       Model |  3684.21053     4  921.052633            Prob > F      =  0.0000
    Residual |  885.877088    31  28.5766803            R-squared     =  0.8062
-------------+------------------------------           Adj R-squared =  0.7811
       Total |  4570.08762    35  130.573932            Root MSE      =  5.3457

------------------------------------------------------------------------------
povertyrate_2011 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------------+------------------------------------------------------------
     birthrate2010 |   .6510034    .271712     2.40   0.023     .0968431    1.205164
inflation_rate_201112 |  -1.264356   .2806831    -4.50   0.000    -1.836813   -.6918994
           avg_mpce |  -.0160157   .0022361    -7.16   0.000    -.0205763   -.0114552
ruralunemployment200910 |   .0882685   .0227547     3.88   0.001     .0418599    .1346771
              _cons |   48.00362   9.624729     4.99   0.000     28.37386    67.63339
------------------------------------------------------------------------------
```

The model estimated is statistically significant with an adjusted R-squared value of 0.7811. Hence it has a very high explanatory power.

Average_MPCE$_{2009\text{-}10}$ & Inflation Rate$_{2011}$ both have a negative impact on Poverty Rate, whereas

Birth_Rate$_{2010}$ & Rural_Unemployment$_{2009\text{-}10}$ both have a positive impact on Poverty Rate

All the results are as per expectation except for one surprising result, negative correlation of inflation rate with poverty rate which we expected to be positive, however inflation reduces real wages, then employment should rise, creating more income-earning opportunities for workers. Therefore, the employment effect of inflation (creating more jobs because of lower labor costs) can outweigh the real-wage effect (lower income) on poverty. This is likely to be the case, as the inflation (real wage) elasticity of poverty is found to be significantly less than the output (employment) elasticity of poverty (UN Report, 2010). Furthermore, most of the poor are net debtors and inflation reduces the real value of their debt. So this way inflation may have a negative correlation with poverty

Another possible explanation is that in the low income countries, along with other countries, poverty is declining (Figure 1). More people can afford goods and services that were previously out of their reach. So demand for essential commodities is increasing at a

higher rate in low income countries. Thus, reduced levels of poverty are pushing demands for consumer goods and services up. As a result, the prices increase and hence, a negative relationship between poverty level and inflation maybe observed in the case of low income countries.

Another interesting thing to note is the lack of "**Growth of GDP**" and "**Literacy Rate**" in our model. This may suggest that overall economic growth does not necessarily lead to reduction in poverty in our country rather a more focused approach towards this sector is required

Absence of **Literacy Rate** goes on to show that simply improving the education may not be the ideal solution to increase income earning opportunities as job opportunities may also be very limited that require a high skilled labour, and thats why most of the employee demand in India is of unskilled labour

Two of the variables with highest coefficients are **Birth rate & Unemployment rate (Rural)** and both are positively correlated with Poverty Rate. Thus Government policies should be more focused on reducing Birth rate and unemployment rate to help reduce Poverty Rate more effectively

CONCLUSION & FURTHER TESTING/EXAMINATION

In conclusion we find birth rate and unemployment rate (rural specifically) are positively correlated with Poverty Rate, and inflation rate and monthly per capita expenditure are negatively related with Poverty Rate.

Robustness tests can be carried out to test whether our results would still hold true if a different estimation technique than OLS is used or if a different measure of poverty is used.

Tests must also be carried out to mitigate endogeneity concerns because it is possible that an omitted variable drives both the dependent and independent variables or it's possible that the dependent variable also affects the distribution of independent variables. So we must apply tests to prove the exogenous variation of the independent variables.

We must also check for selection bias. Selection bias occurs when the procedures used to select subjects and other factors that influence participation in the study produce a result that is different from what would have been obtained if all members of the target population were included in the study. Since data contains only the variables from year 2009-2011, it is possible that if different years were chosen we might have got different results.

We must check whether there is any information bias since systematic errors in the measurement of independent variables is possible.