Phase 4: Final Paper

Group No.: 25

Mukul Nayak (1906323), Ajit Singh Yadav (180030001), Shivam Tiwari (180030029)

Abstract:

The purpose of this study is to examine the effect of high school education (grades IX-XII) on poverty. The Gross Enrolment Ratio for Classes IX-XII is the education variable. The poverty Rate during 2011-12 is the dependent variable. Other independent variables used include labor force participation rate, literacy rate, unemployment rate, percentage of the population living in urban areas and per capita net state domestic product. We gathered data for this study from 2011 to 2012.

To begin, we created a simple regression model with only a primary independent variable and a dependent variable to estimate the impact of education on poverty in the absence of other variables. Then we created a multiple regression model that included all other explanatory variables and discussed their effects and significance. Then we created model 3 by removing statistically insignificant variables from model 2. We used F-statistics and t-statistics to determine whether or not the variables were statistically significant.

We discovered a significant negative relationship between gross enrolment ratio i.e. education variable and poverty through this research.

Topic: The Effect of Education on Poverty

Introduction & Motivation

Education is the only way to live a better life as it directly affects the way people think and react to real-world situations. Access to high-quality education is also known to reduce poverty. Various factors such as economic development, maternal and newborn mortality, and HIV/AIDS can also be linked to education. People have long relied on education to get more excellent pay in the market. It allows us to expand our knowledge, develop new talents, progress as a person, and obtain helpful experience.

In this research, different regression models will be used on the state-wise cross-sectional data of India from 2011-12 to understand the relationship between enrolment in high school and poverty along with some other explanatory variables.

Literature Review

Citak and Duffy (2020) studied the effects of education on poverty in Turkey. They used a cross-sectional study to analyze the two-way causality between the household head's education level and poverty in Turkey. The researchers used an Instrumental Variable (IV) estimation technique, two-stage least squares (2SLS) regression to analyze the effects of education on poverty in Turkey. They then compared the effects of two education reforms in 1961 and 1997. They then identified the causal relationship between education and poverty.

The researchers found that the educational reforms in 1961 and 1997 increased the number of years of schooling by about 20 percent and 9 percent, respectively. They also found that these reforms led to an increase in household income by about 7%. They concluded that the number of schooling years increased due to the reforms, which resulted in a higher household income.

Lupeja and Gubo (2017) examined the contribution of knowledge and skills acquired from secondary education to self-employment among graduates in the Mvomeero District of Tanzania. Through a quantitative method, 400 individuals were interviewed. The study was conducted through a systematic interviewing process. A self-administered questionnaire was then used to collect opinions on the effects of knowledge and skills acquired from secondary education on poverty reduction. The results indicated that secondary education could contribute to reducing poverty.

Janjua and Kamal (2011) analyzed the data collected from 40 developing countries from 1999 to 2007. It estimated the coefficients by applying the random effect generalized least squares (GLS) technique. It found that income growth is associated with a positive effect on reducing poverty but does not play a significant role in reducing poverty. The study led to three conclusions, i.e., during the observed period, per capita income growth played a moderate role in poverty reduction in the selected countries; only in nations with higher per capita incomes did income inequality have a more significant influence on poverty alleviation. Finally, secondary education has emerged as the most significant contribution to poverty reduction.

Njong (2010) analyzed the effects of different levels of education on employed individuals as determinants of poverty in Cameroon. The data for this research came from a 2001 household survey in Cameroon. A sample-selectivity adjusted logistic regression model was used to analyze the data. The results indicated that the probability of being poor was related to the employed individual’s level of education and experience. The results show that improving experience and education reduces the probability of being poor of the employed individual. Regarding gender, the study concludes that men’s education levels help reduce poverty more than women’s.

The literature review leads us to various conclusions like education may increase an individual’s income by increasing productivity hence significantly reducing poverty. Aside from being beneficial to the individual, education also has a wide range of externalities that can improve the lives of poor people. For instance, it can help lower infant mortality, improve parental education, reduce health risks, reduced stunting, and reduced violence at home and in society. The effects of education on poverty can vary depending on the region and the level of education. This suggests that studying the link between education and poverty is essential. There is also a need to study the various factors that affect education development in different regions. It is evident that we must first understand what causes it to reduce poverty effectively. This study plans to do the same.

Research Question:

I. What is the Effect of Education on Poverty?
II. How is Poverty related to an increase in High School's Gross Enrolment Ratio?

This paper will discuss the effect of education on poverty. As education increases, people acquire various skills that help them land a good job, and their income sources increase; hence poverty decreases. Hence there is a causal relationship between education and poverty.

Hypothesis: Education has a negative impact on poverty rates; therefore, as education rates increase poverty rate decreases.

In this study, we develop a regression model with education level as our independent variable and examine its causality upon poverty rate. This study is to research on the education-poverty relationship and try to check whether the hypothesis is correct or not.

Methodology

Data

Cross Section type data of 32 States and UTs of India is taken for the year 2011-12. The dependent variable is the state-wise Poverty Rate (*pov*) of India and the primary independent variable used is the state-wise Gross Enrolment Ratio (*ger*) for classes IX-XII.

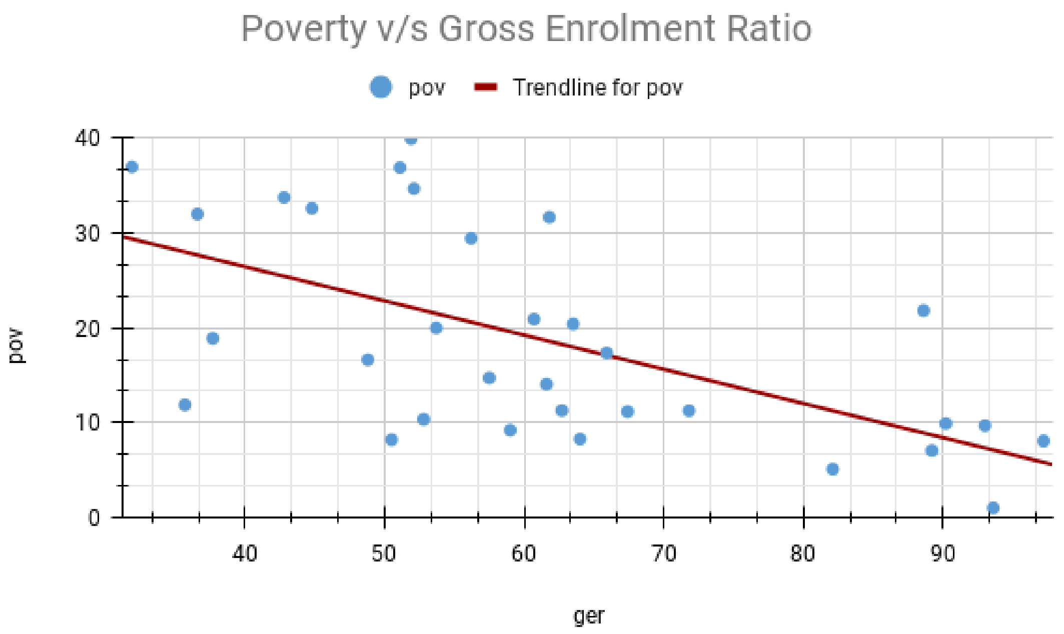

Figure 1 – Scatter Plot of *pov* v/s *ger*

A few more independent variables are included in the multiple linear regression model to examine the ceteris paribus effect of gross enrolment ratio on the poverty rates of Indian states. The other independent variables are Literacy Rate, Labour Force Participation Rate, Unemployment Rate, Per Capita Net State Domestic Product and Percentage of Population in Urban Area.

Table 1 – Variables Description

Sl No	***Name***	***Description***	***Source***
1	*pov*	Poverty Rate (%) 2011-12 (Based on MRP Consumption)	Planning Commission, National Sample Survey Organization (NSSO), Government of India.
2	*ger*	Gross Enrolment Ratio (Classes IX-XII)	Statistics of School Education 2011-12
3	*lit*	Literacy Rate (%)	Census 2011
4	*lfpr*	Labour Force Participation Rate (per 1000)	Key Indicators of Employment and Unemployment in India, NSS 68th Round (2011-12)
5	*unemp*	Unemployment Rate (per 1000)	Key Indicators of Employment and Unemployment in India, NSS 68th Round (2011-12)
6	*nsdp*	Per Capita Net State Domestic Product in ₹ , 2011-12 (At Current Prices) (Base : 2011-12)	National Statistical Office (NSO)
7	*urb*	Percentage of Population in Urban Area (%)	Handbook of Statistics on Indian States, Reserve Bank of India

The descriptive statistics for each variable are shown in the table below.

Table 2 – Descriptive Statistics

Variable	***Sample Size***	***Mean***	***Median***	***Standard Deviation***	***Minimum***	***Maximum***
pov	32.000	18.592	15.670	11.086	1.000	39.930
ger	32.000	61.763	59.850	18.256	31.900	97.200
lit	32.000	77.176	76.600	8.448	61.800	94.000
lfpr	32.000	408.969	404.000	52.635	283.000	526.000
unemp	32.000	35.469	25.000	35.144	5.000	177.000
nsdp	32.000	83101.656	71270.000	50095.139	21750.000	259444.000
urb	32.000	36.228	29.716	20.984	10.036	97.504

Methodology

To begin, a simple regression model will be created to check the hypothesis and to estimate the ceteris paribus impact of education on poverty.

$$pov = \beta_0 + \beta_1(ger) \pm \varepsilon$$

Then model 2 will be developed incorporating other explanatory variables.

$$pov = \beta_0 + \beta_1(ger) + \beta_2(lit) + \beta_3(lfpr) + \beta_4(unemp) + \beta_5(nsdp) + \beta_6(urb) \pm \varepsilon$$

In this model statistical significance of variables will be checked using t-statistics and p-values.

Now, the final multiple regression model will be created by omitting statistically insignificant variables from the model 2. And finally using F-tests it would we be checked whether those eliminated variables are jointly significant or not.

We will analyze the education-poverty relationship in this study and see if our hypothesis is valid or not.

Empirical Results:

Model 1 – Simple Regression Model

In this simple regression model, dependent variable is *pov* and independent variable is *ger*.

$$pov = \beta_0 + \beta_1(ger) \pm \varepsilon$$

OLS Regression Results						
Dep. Variable:		*pov*	R-squared:			0.353
Model:		OLS	Adj. R-squared:			0.331
Method:		Least Squares	F-statistic:			16.36
No. Observations:		32	Prob (F-statistic):			0.000338
Df Residuals:		30	Log-Likelihood:			114.92
Df Model:		1	AIC:			233.8
Covariance Type:		nonrobust	BIC:			236.8
	coef	**std err**	**t**	**p>|t|**	**[0.025**	**0.975]**
const	40.8715	5.737	7.125	0.000	29.156	52.588
ger	-0.3607	0.089	-4.045	0.000	-0.543	-0.179
Omnibus:		1.738	Durbin-Watson:			1.855
Prob(Omnibus):		0.419	Jarque-Bera (JB):			1.206
Skew:		0.206	Prob(JB):			0.547
Kurtosis:		2.143	Cond. No.			230

From the above OLS regression results, the estimated equation is as follows:

$$pov = 40.8715 - 0.3607(ger) \pm 0.089$$

The coefficient term is -0.3607 for the primary independent variable *ger*, this negative coefficient term validates the hypothesis that education has a negative impact on poverty rates, it indicates if *ger* increases by 1 unit then *pov* decreases by 0.3607 units.

The t value for *ger* is -4.045. The p-value is 0.000 which means it is <0.001 (but not exactly 0) hence we can say that *ger* is statistically significant at the 1% significance level. The R-squared for this model is 0.353 which means 35.3% variation in *pov* can be explained by *ger*.

Model 2 – Multiple Regression Model (incorporating other explanatory variables)

To further enhance the Model 1, other explanatory variables like *lit*, *lfpr*, *unemp*, *nsdp* and *urb* are also taken into account.

$$pov = \beta_0 + \beta_1(ger) + \beta_2(lit) + \beta_3(lfpr) + \beta_4(unemp) + \beta_5(nsdp) + \beta_6(urb) \pm \varepsilon$$

OLS Regression Results						
Dep. Variable:			*pov*	R-squared:		0.533
Model:			OLS	Adj. R-squared:		0.421
Method:			Least Squares	F-statistic:		4.752
No. Observations:			32	Prob (F-statistic):		0.00234
Df Residuals:			25	Log-Likelihood:		-109.70
Df Model:			6	AIC:		233.4
Covariance Type:			nonrobust	BIC:		243.7
	coef	**std err**	**t**	**p>|t|**	**[0.025**	**0.975]**
const	67.5886	18.262	3.701	0.001	29.978	105.200
ger	-0.2435	0.126	-1.926	0.066	-0.504	0.017
lit	-0.1023	0.333	-0.307	0.761	-0.788	0.583
lfpr	-0.0518	0.035	-1.501	0.146	-0.123	0.019
unemp	-0.0329	0.051	-0.649	0.523	-0.137	0.072
nsdp	-8.546e-05	4.62e-05	-1.849	0.076	-0.000	9.72e-06
urb	0.0937	0.122	0.765	0.451	-0.159	0.346
Omnibus:			1.029	Durbin-Watson:		2.071
Prob(Omnibus):			0.598	Jarque-Bera (JB):		0.649
Skew:			0.348	Prob(JB):		0.723
Kurtosis:			2.949	Cond. No.		1.18e+06

From the above OLS regression results, the estimated equation is as follows:

$$pov = 67.5886 - 0.2435(ger) - 0.1023(lit) - 0.0518(lfpr) - 0.0329(unemp) \\ - 0.00008546(nsdp) + 0.0937(urb) \pm \varepsilon$$

The R-squared value for this model is 0.533 which means 53.3% variation in *pov* can be explained collectively by the independent variables around its mean.

The coefficient term for the primary independent variable *ger* is -0.2435, this negative coefficient term still validates the hypothesis that education has a negative impact on poverty rates and if *ger* increases by 1% then *pov* decreases by 0.2435%. Moreover, the t-statistic for *ger* is -1.926 and the p-value is 0.066 which means it is statistically significant at close to 5% level of significance.

The p-values for *lit*, *unemp* and *urb* are 0.761, 0.523 and 0.451 respectively, which are not statistically significant even at 10% level, hence these statistically insignificant variables can be removed in further models.

The t-statistic for *lfpr* is -1.501 and the p-value is 0.146 which means it is statistically significant at 15% level of significance.

The coefficient term for *nsdp* is -0.00008546 which is very low but it has t-stats of -1.849 and the p-value is 0.076, therefore it is statistically significant at the 10% significance level. Hence, it will not be eliminated.

Model 3 – Multiple Regression Model (omitting statistically insignificant variables)

Statistically insignificant variables like *lit, unemp, and urb* are eliminated and new multiple regression is being performed in this model.

$$pov = \beta_0 + \beta_1(ger) + \beta_2(lfpr) + \beta_3(nsdp) \pm \varepsilon$$

OLS Regression Results			
Dep. Variable:	*pov*	R-squared:	0.509
Model:	OLS	Adj. R-squared:	0.456
Method:	Least Squares	F-statistic:	9.668
No. Observations:	32	Prob (F-statistic):	0.000152
Df Residuals:	28	Log-Likelihood:	-110.51
Df Model:	3	AIC:	229.0
Covariance Type:	nonrobust	BIC:	234.9

	coef	**std err**	**t**	**p>|t|**	**[0.025**	**0.975]**
const	65.1242	12.252	5.315	0.000	40.027	90.222
ger	-0.2337	0.101	-2.320	0.028	-0.440	-0.027
lfpr	-0.0644	0.028	-2.304	0.029	-0.122	-0.007
nsdp	-6.908e-05	3.67e-05	-1.885	0.070	-0.000	5.99e-06

Omnibus:	2.699	Durbin-Watson:	1.994
Prob(Omnibus):	0.259	Jarque-Bera (JB):	1.863
Skew:	0.589	Prob(JB):	0.394
Kurtosis:	3.101	Cond. No.	8.19e+05

From the above OLS regression results, the estimated equation is as follows:

$$pov = 65.1242 - 0.2337(ger) - 0.0644(lfpr) - 0.00006908(nsdp) \pm \varepsilon$$

The R-squared value for this model is 0.509 which means 50.9% variation in *pov* can be explained collectively by the independent variables *ger*, *lfpr* and *nsdp* around its mean.

The coefficient term for the primary independent variable *ger* is -0.2337, this negative coefficient term still validates the hypothesis that education has a negative impact on poverty rates and if *ger* increases by 1% then *pov* decreases by 0.2337%. Moreover, it has t-stats of -2.320 and a p-val of 0.028 which means it is statistically significant at 5% significance level.

The coefficient term for *lfpr* is -0.0644 which means if *lfpr* increases by 1% then *pov* decreases by 0.0644% and it has t-stats of -2.304 and a p-val of 0.029 which means it is also statistically significant at 5% significance level.

The t-statistic of *nsdp* is -1.885 and a p-val of 0.070 which means it is statistically significant at the 10% significance level.

Table 3 – Summary of Models 1,2 & 3

Dependent Variable : *pov*				
		Model 1	**Model 2**	**Model 3**
Independent Variables	*ger*	-0.3607*** (0.089)	-0.2435** (0.126)	-0.2337** (0.101)
	lit	-	-0.1023 (0.333)	-
	lfpr	-	-0.0518 (0.035)	-0.0644** (0.028)
	unemp	-	-0.0329 (0.051)	-
	nsdp	-	-0.00008546* (0.0000462)	-0.00006908* (0.0000367)
	urb	-	0.0937 (0.122)	-
Intercept		40.8715 (5.737)	67.5886 (18.262)	65.1242 (12.252)
Sample Size		32	32	32
R-squared		0.353	0.533	0.509
Adj. R-squared		0.331	0.421	0.456
F-statistic		16.36	4.752	9.668
Prob(F-statistic)		0.000338	0.00234	0.000152
Durbin-Watson		1.855	2.071	1.994
Jarque-Bera (JB)		1.206	0.649	1.863
Prob(JB)		0.547	0.723	0.394

*Significant at *10%, **5%, ***1%*

F-Test:

In Model 2 it was concluded that *lit, unemp* and *urb* were not statistically significant when taken individually and hence they were eliminated in Model 3. Now using F-test it will be examined whether they are jointly significant or not.

For this test Model 3 will be Restricted Model and Model 2 will be Unrestricted Model.

Restricted Model:

$$pov = 65.1242 - 0.2337(ger) - 0.0644(lfpr) - 0.00006908(nsdp) \pm \varepsilon$$

Unrestricted Model:

$$pov = 67.5886 - 0.2435(ger) - 0.1023(lit) - 0.0518(lfpr) - 0.0329(unemp) - 0.00008546(nsdp) + 0.0937(urb) \pm \varepsilon$$

Hypothesis:

$$H_0 : \beta_2 = \beta_4 = \beta_6 = 0$$

$$H_1 : \beta_2 = \beta_4 = \beta_6 \neq 0$$

On Solving:

$$F = \frac{(R_{um}^2 - R_{rm}^2)/q}{(1 - R_{um}^2)/(n-k-1)}$$

$$F = \frac{(0.533 - 0.509)/3}{(1 - 0.533)/(32 - 6 - 1)}$$

$$F = \frac{0.008}{0.467/25}$$

$$F = 0.428$$

From the F-Distribution tables the critical values are known i.e.

$F_{0.001,3,25} = 7.45$,

$F_{0.010,3,25} = 4.68$,

$F_{0.025,3,25} = 3.69$,

$F_{0.050,3,25} = 2.99$,

$F_{0.100,3,25} = 2.32$

Now,

$F(=0.428) < F_{0.001,3,25}(=7.45)$, $F(=0.428) < F_{0.010,3,25}(=4.68)$, $F(=0.428) < F_{0.025,3,25}(=3.69)$, $F(=0.428) < F_{0.050,3,25}(=2.99)$ and $F(=0.428) < F_{0.100,3,25}(=2.32)$, hence *lit, unemp* and *urb* are not jointly significant at any of the 0.1%, 1%, 2.5%, 5% and 10% of significance level.

It can be seen that we fail to reject null hypothesis at every significance level discussed above, hence it can be concluded that *lit, unemp* and *urb* are not jointly significant (or jointly insignificant).

Conclusion:

The original hypothesis is still valid after testing the significance of the various independent variables in different models. Each model shows a significant negative relationship between poverty rate and gross enrolment ratio.

In Model 1, simple linear regression was used to test if *ger* significantly affected *pov*. The overall regression was statistically significant with R-squared of 0.353 and *ger* was statistically significant at the 1% significance level.

In Model 2, simple linear regression with all other explanatory variable was used. R-squared value for this model is 0.533. It was concluded that *lit, unemp* and *urb* were statistically insignificant.

In Model 3, the statistically insignificant variables were eliminated and R-squared was found to be 0.509. The R-squared value slightly decreased from Model 2 but Adj. R-squared was increased from 0.421 to 0.456 which means this model is better.

In F-Test it can be seen that we fail to reject null hypothesis at every significance level, hence model 3 is better than model 2.

The coefficient term for the primary independent variable *ger* is -0.3607 and -0.2435 in model 1 and model 2 respectively. And in model 3 the coefficient term for *ger* is -0.2337 which means if *ger* increases by 1% then *pov* decreases by 0.2337%. Moreover, it has t-stats of -2.320 and a p-val of 0.028 which means it is statistically significant at 5% significance level.

From model 2 and F-Test we can conclude that *lit, unemp* and *urb* are statistically insignificant when considered jointly or independently.

Throughout the study, only *ger, lfpr* and *nsdp* were consistently statistically significant.

We cannot observe the influence of education on poverty in other countries or on grade school education since the data in the study solely focus on India and only examine high school education. The findings of this analysis show that additional research in this area is required. Various other significant independent variables can also be added to enhance the Model 3 and get a better R-squared value. Cross-country analysis can also help to further the investigation.

Bibliography

- Citak, Ferhat & Duffy, Patricia. (2020). The Causal Effect of Education on Poverty: evidence from Turkey. 11. 251-265.
- Lupeja, T.L., & Gubo, Q. (2017). Secondary Education Attainment and its Role in Poverty Reduction: Views of Graduates Working in Informal Sector in Rural Tanzania. *Journal of Education and Practice, 8*, 140-149.
- Janjua, P. and Kamal, U. (2011). The Role of Education and Income in Poverty Alleviation: A Cross-Country Analysis. *The Lahore Journal of Economics.* Vol. 16, No. 1, p. 143-172
- Njong, A.M. (2010). The effects of educational attainment on poverty reduction in Cameroon. *International Journal of Educational Administration and Policy Studies, 2*, 001-008.

Appendix:

Screenshot of Data Sheet

	States and UTs	pov	ger	lit	lfpr	unemp	nsdp	urb
2	Andaman and Nicobar Islands	1	93.6	86.63	449	65	89100	37.53
3	Andhra Pradesh	9.2	59	67.02	479	20	69000	33.36
4	Arunachal Pradesh	34.67	52.1	65.38	376	22	73540	22.90
5	Assam	31.98	36.6	72.19	358	46	41142	14.10
6	Bihar	33.74	42.8	61.8	283	34	21750	11.30
7	Chandigarh	21.81	88.6	86.05	376	60	158967	97.25
8	Chhattisgarh	39.93	51.9	70.28	469	14	55177	23.24
9	Delhi	9.91	90.2	86.21	350	38	185001	97.50
10	Goa	5.09	82.1	88.7	376	49	259444	62.17
11	Gujarat	16.63	48.8	78.03	424	5	87481	42.60
12	Haryana	11.16	67.4	75.55	354	29	106085	34.88
13	Himachal Pradesh	8.06	97.2	82.8	526	13	87721	10.04
14	Jammu and Kashmir	10.35	52.8	67.16	403	34	53173	27.37
15	Jharkhand	36.96	31.9	66.41	360	26	41254	24.05
16	Karnataka	20.91	60.7	75.36	430	16	90263	38.67
17	Kerala	7.05	89.2	94	403	66	97912	47.70
18	Madhya Pradesh	31.65	61.8	69.32	388	9	38497	27.63
19	Maharashtra	17.35	65.9	82.34	437	13	99597	45.22
20	Manipur	36.89	51.1	76.94	386	37	39762	29.20
21	Meghalaya	11.87	35.7	74.43	438	8	60013	20.05
22	Mizoram	20.4	63.5	91.33	446	32	57654	52.14
23	Nagaland	18.88	37.7	79.55	445	177	53010	28.85
24	Odisha	32.59	44.8	72.87	422	24	48499	16.69
25	Puducherry	9.69	93	85.85	362	21	119649	68.35
26	Punjab	8.26	64	75.84	401	22	85577	37.48
27	Rajasthan	14.71	57.5	66.11	405	12	57192	24.87
28	Sikkim	8.19	50.5	81.42	525	12	158667	25.20