Public Education Expenditures and their Effects on State Income

Alexis Andrews

Cathy Campbell

Sourabh Chatterji

Abstract:

Education is considered to be one of the most vital factors in determining the success of an economy. This paper examines the connection between public expenditures toward post secondary education and state income levels. In a more complex multiple regression model we take into account unemployment rates, number of metro areas, and poverty rates within each state. In both the simple and the multiple regression models we found that public expenditures for post-secondary education are statistically significant in its effect on state income.

Introduction

In this paper we studied the connection between investment in education and state income. As college students, we were interested in seeing how the funding put towards education in our state actually affects the economy. Also, the results of this research could have some important policy implications as far as how to target government funding within the states. We hope to find through our research if increasing expenditures in public education would have positive, long-lasting effects on the economy.

We expect education expenditures to have a positive correlation with state income. An educated labor force is presumably more productive for educated workers are better able to utilize existing capital. This would allow firms to produce more without having to increase costs, so income would continue to increase. Educated workers are also more innovative, and the technological advances they provide lead to economic growth.

Firms also tend to build headquarters near top rated universities, so they can more easily recruit qualified students as workers. As more firms continue to move to states with these schools, there is an increase in overall output in the states. Also, since several firms are competing for the same pool of limited applicants, they have to offer competitive wages in order to ensure they hire the most qualified workers. This migration of firms brings jobs to the area, possibly making way for those previously unemployed residents to work again, also shifting the average state income up. In short, public investments into post-secondary education bring firms to the state, who in turn produce more using the educated laborers.

Literature Review

Eid (2012) looks at the relation between research and development (R&D) investment and economic productivity growth in 17 high-income OECD countries over the time period 1981-2006. Eid looked at both R&D done through higher education and R&D funded through private companies and the government. He found that R&D performed by universities actually had higher returns in productivity than R&D done in the private and government sector. The long-run propensity of productivity growth was found to be more significant with higher education R&D than the private or government R&D as well. This relates to our paper because we are looking at how the investment in higher education influences income; this article considers how higher education influences productivity, and we know by the Solow model that where there is more productivity, there is more growth, meaning higher incomes (Eid 53-68).

Breton's (2005) paper on the role of education in economic growth discusses the very concept we are exploring in this paper. Breton used cross-country data to compare 2005 national incomes and connects these figures to noted expenditure towards post-secondary education. It is explained that education is often a "limiting factor" that determines economic growth and then quantifies the significance of the effects of education on the economy. Breton then further explains the data found and how it should affect government policy on education. The results show that education has direct and indirect effects on national income. On average, an educated workforce has a higher marginal productivity, thus increasing national income. Indirect effects are noted by increased marginal productivity of other workers. The empirical results indicate that investment in education has a high marginal return, at over 10% in highly educated countries. Breton argues that public investment in education is vital to experiencing these high returns. Breton's findings are right along with our hypothesis that an increase in public expenditures towards post-secondary education lead to an increase in gross income.

Baldwin (2011) studies the effects of state educational investment on economic growth through changes in gross state product. The study was conducted by collecting longitudinal data on 48 states within the continental United States. The data was gathered from 1988 to 2005 and data from other intervening variables was collected for a period of 16 years. Educational expenditures were grouped into four year averages to offset for short term economic downturns. The study relied on the change in the growth rate of Gross state product to measure the effects of educational investment, amongst other variables. A total of nine variables were studied to measure their effects on the growth rate of gross state product. All the independent variables in the model (except for high-school attainment) predicted the growth of gross state product from 1997 to 2005.

The concept of education expenditure leading to higher income has been studied extensively. Our research would provide additional insight into the concept by looking at the dynamics of education investment on a much smaller scale. Most of the previous work has been done on the national level, while ours examines the effects in different sections of the United States, i.e., the individual states. This approach could lead to some interesting results, since certain economic factors have more effects on an aggregate level as opposed to a smaller scale.

Data

For our study, we looked at a total of six variables. We chose these variables based on the rationale that they would have some effect on growth of state income (y). We chose our first variable, educational expenditures per full time student (x_1), to determine if our hypothesis was backed up by solid data. We were interested in determining whether state's that spent more money per pupil on higher education saw a return on that investment in the form of increased state income (y). However, to get a more accurate picture of the relationship between our main variable and growth in state income, we had to establish other variables that might affect the growth as well. As a result, we found five other variables that we decided might also have a significant relationship with our dependent variable (growth in state income). The second variable that was included was the unemployment rate within each state (x_2). A high unemployment rate could be a cause of a struggling economy, and it would have a direct impact on state per capita income. The third variable in our study was the number of metro areas within each state (x_3). State's that contain a large number of metro areas usually implies the state has a high population, more jobs, and more growth. Naturally, this would mean a significant relationship with per capita income within the state. The fourth variable was poverty rates within each state (x_4). A large poverty rate can imply that a state has a struggling economy, and it can also indicate that the state has an unequal distribution of income. Our last variable was the percent of adults ages 25-34 that have a post-secondary degree within each state (x_5). This variable indicates the level of human capital within each state. Also, a high percentage of adults with postsecondary degrees can imply that a state has many skilled workers who typically earn more than unskilled workers.

We did share a few variables with the study done in Baldwin (2011). The common variables were x_1 (higher education expenditures per capita in Baldwin (2011)) and x_5 (% high school attainment in Baldwin (2011)). However, while we did share certain independent variables, there were fundamental differences in how the data for each variable was collected and implemented. In Baldwin (2011), the study uses data from 1982 to 1998 for higher education expenditures per capita. For our study, we looked at data from only one fiscal year (2009-2010). Also in Baldwin (2011), when collecting data for % High school Attainment, the study assumes a seven-year lag between enrolling in higher education to presumed effect on state GDP. In our study, we did not use a lag due to our assumption that rates were remaining largely static.

Our data was collected from the Census Bureau, the US Department of Education, the Brookings Institute, and the Bureau of Labor Statistics. The Census Bureau, in particular, was helpful for finding a large part of our data. It contains vast amounts of data related to education. The Census Bureau contained the data for state per capita income (y) in 2009, educational appropriations per full time student (x_1) for fiscal year 2009-2010, and poverty rates (x_4) from 2009-2010. The Brookings Institute contained the data on the number of metro areas (x_3) from 2009. The Bureau of Labor Statistics contained the data on state unemployment rates (x_2) from 2009. Finally, the US Department of Education contained the data on the percent of adults ages 25-34 that had post-secondary degrees (x_5) from 2010. Table 1 shows the data, years, and sources. Table 2 shows the summary statistics of the variables.

Table 1:

Variable	Year(s)	Source
Average State income (y)	2009	Census Bureau
Educational appropriations per full-time student (x_1)	2009- 2010	Census Bureau
Unemployment Rate (x_2)	2009	U.S. Bureau of Labor Statistics
Number of Metro Areas (x_3)	2010	Brookings Institute
Poverty Rates (x_4)	2009- 2010	Census Bureau
% of Adults Ages 25-34 with Post-secondary Degree (x_5)	2010	U.S. Department of Education

Table 2:

Variable	Obs	Mean	Std. Dev.	Min	Max
eduexp	50	6611.54	2082.842	3073	14940
avginc	50	50274.18	8450.441	36646	69272
unemploy	50	8.432	1.954528	4.1	13.5
metros	50	8.4	5.989787	1	26
povrate	50	13.876	3.237885	7.1	22.8
post2nddeg~t	50	38.898	6.457253	28.4	54.3

To ensure unbiasedness, we need to test to see if our data follows the five Gauss-Markov Assumptions. Assumption 1 requires that the model is linear in parameters. All of our coefficients are linear in parameters. Assumption 2 requires that there is a random sample of n observations. In our model, we have data from a given time from 50 states. This satisfies this assumption. Assumption 3 requires that there be no perfect collinearity amongst the independent variables. In our model, some variables might have some correlation; however, there is no variable that is perfectly correlated with another. Assumption 4 requires that the error has an expected value of zero given any x_k. To try to satisfy this assumption, we added more variables to our model to attempt to ensure that no important variables are omitted. Assumption 5 does not have to do with establishing unbiasedness, but has to do with variance. It says that the error should have some constant variance given any x. It can also be interpreted that the variance of y will not change based on any x. As with assumption 4, we tried to establish variables that could satisfy this assumption and not exhibit homoskedasticity.

Table 3 shows the correlation coefficients, and the following figures show them graphically.

Table 3:

	eduexp	unemploy	metros	povrate	post2n~t
eduexp	1.0000				
unemploy	-0.1036	1.0000			
metros	-0.1698	0.4668	1.0000		
povrate	-0.0525	0.3984	0.3403	1.0000	
post2nddeg~t	-0.1086	-0.2767	-0.1146	-0.6537	1.0000

In Figure 1, we see a slightly positive correlation between average income and education expenditure per pupil on higher education.

Figure 1:

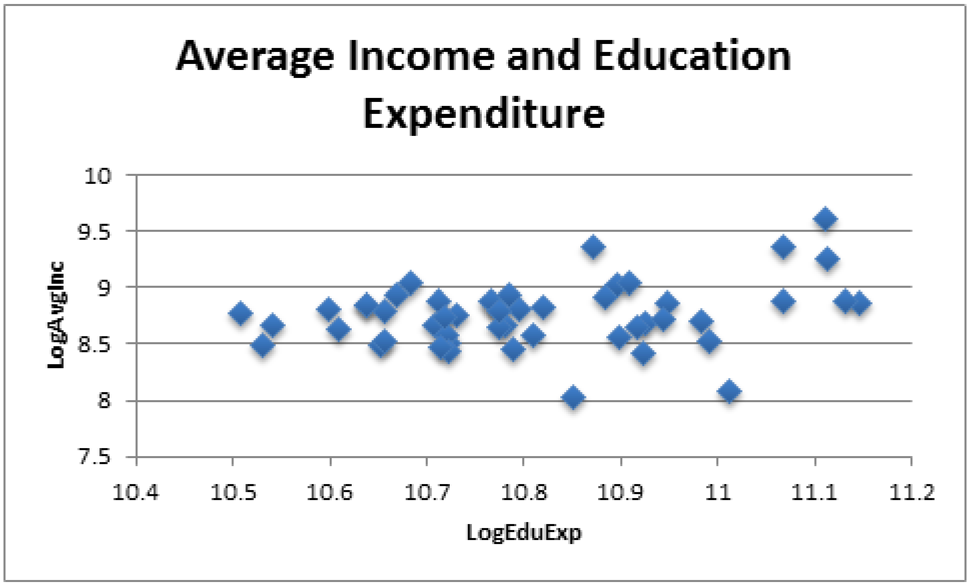


Figure 2 shows a scatter chart with log of average income against log of population density, log of number of metropolises in the state, and log of per-pupil education expenditure on higher education. We see a slightly positive correlation for the positive correlation for the log of population density and log of education expenditure variables and a negative correlation of the log of metropolitan areas variable.

Figure 2:

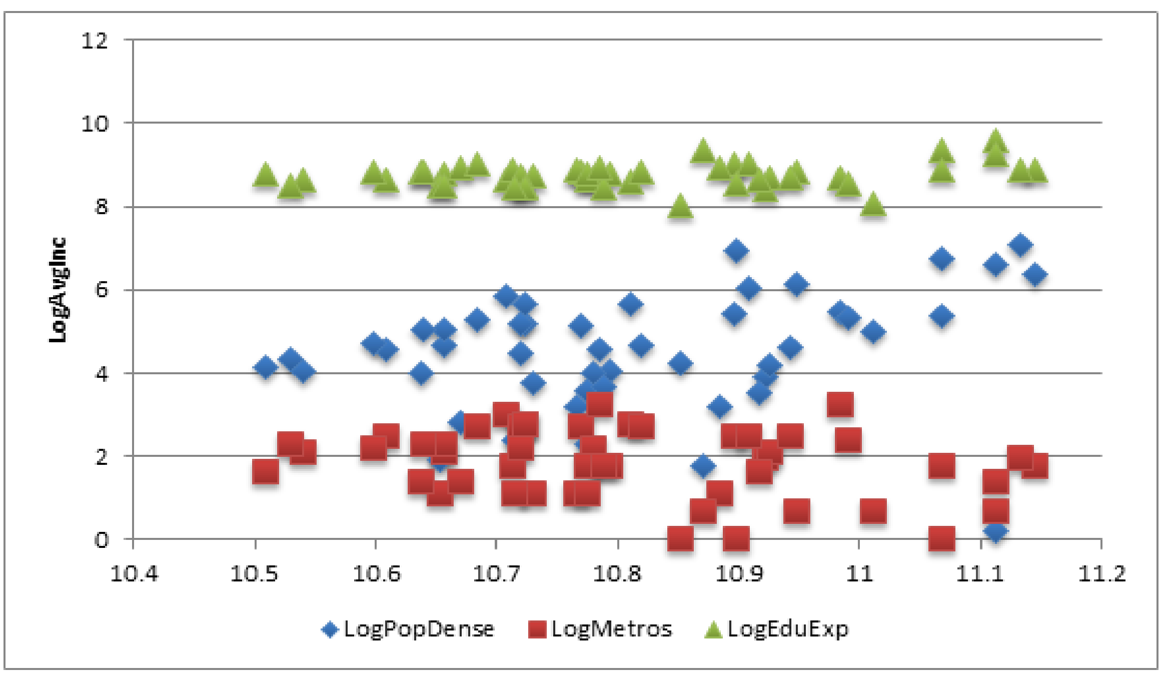
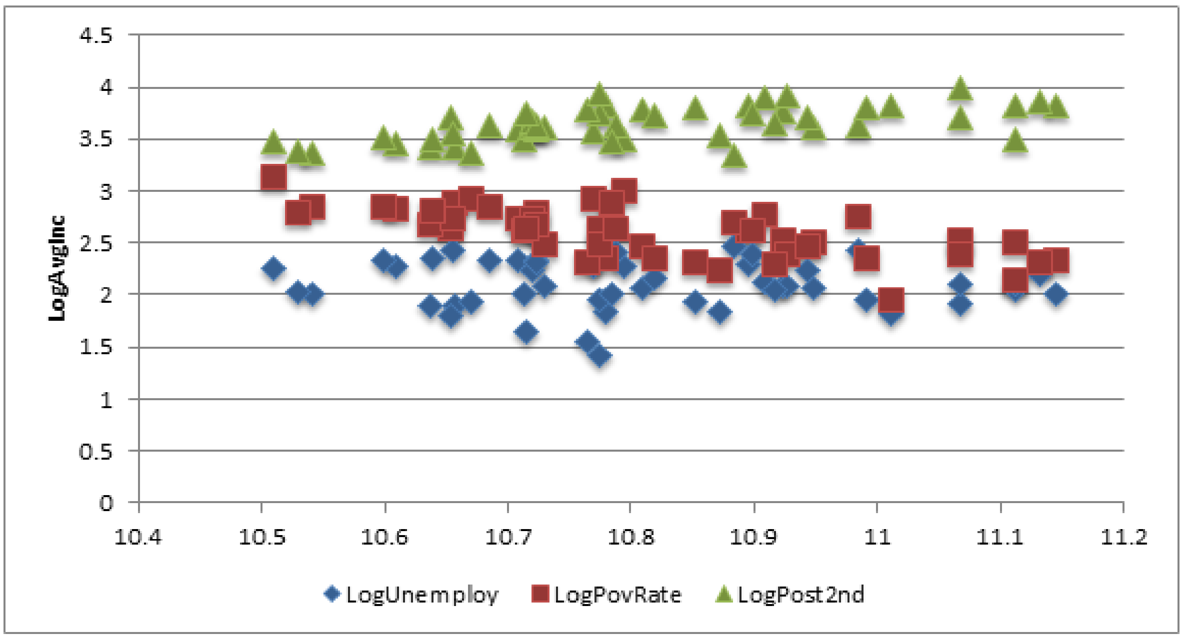


Figure 3 shows a scatter plot with log of average income against log of unemployment in the state, log of the state's poverty rate, and the log of the percent of the population of the state with a post-secondary degree. We see a slightly positive correlation for the log of the percent of the population of the state with a post-secondary degree variable, a slight negative correlation for the log of poverty rate variable, and a neutral log of unemployment variable.

Figure 3:



Results

Simple Linear Regression

Our simple regression model gives us this equation:

logavginc = 9.34 + (0.168)logeduexp + 1.202.

Figure 5 in the appendix shows the Stata regression.

As we predicted, there is a slight positive correlation between median income in a state and the expenditure per pupil of higher education by the same state. The coefficient 0.168 gives the elasticity of average income with respect to education expenditure. This means that with every 1% increase in education expenditure, there is a 0.168% increase in average income.

Multiple Regression

Our multiple regression gives us this model:

logavginc = 8.804 + (0.21)logeduexp + (0.024)unemploy - (0.0258)povrate + (0.0096)post2nddegpercent - (0.025)logmetros + 0.412.

Figure 6 in the appendix shows the Stata regression.

The coefficients of each variable with a log on it tell the elasticity of median income with respect to each variable. For the variables without a log, like unemploy, povrate, and post2nddegpercent gives us a percentage interpretation. For example, when povrate is multiplied by 100, logavginc increases by 2.4%.

We didn't expect to see a negative sign on our logmetros variable. We expected that where there are more metropolitan areas, the income would be higher. Every other variable is about like we expected. This may be explained by the logmetros variable being the least significant of all the variables. The rest of the variables are statistically significant at 1%, with t-values all above 2.693, while logmetros is significant at about 21% according to its p-value.

Unemployment rate and poverty are closely related variable, so we have two more multiple regression models each omitting one of the variables to see how that changes the other variable. Figure 7 and Figure 8 in the appendix show the Stata regressions for each model.

We see interesting changes in the coefficients and statistical significances in each of these models. The variables unemploy and logmetros vary between being significant and not, so the fourth multiple regression took those two variables out and we will test the joint significance with an F-stat test against the first multiple regression model. The State regression for the fourth model is Figure 9 in the appendix.

Figure 10 below succinctly summarizes the coefficients, t-stats (in parentheses), and level of significance for each variable along with number of observations and R^2 terms for each model.

Figure 10:

	Dependent Variable : Log (avg Income)				
Independent Variables	Single Regression	1st Multiple Regression	2nd Multiple Regression	3rd Multiple Regression	4th Multiple Regression
logeduexp	0.167** (2.12)	0.209*** (4.26)	0.21*** (3.67)	0.202*** (3.82)	0.204*** (3.90)
unemploy		0.024*** (2.90)	0.019* (1.98)		
Povrate		-0.026*** (-4.14)		-0.023*** (-3.47)	-0.024**** (-3.93)
Post2nddegpercent		0.01*** (3.30)	0.017*** (6.58)	0.009*** (2.80)	0.009*** (2.83)
logmetros		-0.025 -1.28	-0.051** (-2.33)	-0.005 (-0.24)	
constant	9.343*** (13.49)	8.803*** (18.13)	8.232*** (15.17)	9.029*** (17.45)	9.023*** (17.64)
# of observations	50	50	50	50	50
R^2	0.0858	0.6862	0.564	0.6262	0.6257

(Note *** represents 1% level of significance, ** 5% level of significance, and * 10% level of significance)

For the F-stat test, we will use the first multiple regression for the unrestricted model and the fourth multiple regression test for the restricted model.

$H_0 = B_2 = 0, B_5 = 0$

$H_1 = H_0$ not true.

To find the F-stat we use the formula:

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

For our model we get:

$$F = \frac{(0.6862 - 0.6257)/2}{(1 - 0.6862)/44} = \frac{(0.0605/2)}{(0.3138/44)} = \frac{0.03025}{0.00713182} \approx 4.24$$

The critical value on the 5% Critical Values of the F Distribution with numerator degrees of freedom 2 and denominator degrees of freedom 44 is 3.21. Our F-stat is 4.24 which is greater than the critical value. Therefore, we reject H_0. Unemployment rate and log of metropolitan areas are significant jointly.

Conclusion

Our results show that investing in post-secondary education has significant positive effects on the economy of the United States. For those states that have more education investments, they tend to have higher average incomes. When controlling for other possible variables, the effects of education expenditures become even more significant. This result is what we expected when considering the basic economic principles on the relationship between education and economic progress. Our original hypothesis is correct according to our model, however, some of the additional explanatory variables used in the multiple regressions had effects that we had not expected. Mainly, variables referring to the unemployment rate and the number of metro areas in the state affected the model in the exact opposite of the expected manner. As for the number of metro cities, we predicted it would have a positive significant effect on income; however, in two of the 3 multiple regressions, the variable was not significant. Furthermore, it had a negative impact on the average income.

Upon further consideration, we realized that some of the variables may be related. For example, large metropolitans usually have higher poverty rates. Low-income residents lack resources necessary to find viable employment, namely transportation, education, and a network of professional contacts. Cities have public transit systems to connect them to multiple locations. Despite these seemingly favorable conditions, the wages received by these unskilled workers are lacking, therefore keeping them below the poverty line . This would explain why a large city, despite having an abundance of industry to bring economic productivity to the area, would experience negative effects on the state income. As for the unemployment rate, we expected it to have a negative impact on state income, yet it was actually positive. Why would the number of residents not receiving any income have a positive effect on the state income? This anomaly is perplexing and would require further research and possibly a separate hypothesis to test with a different regression model. When we tested the variables for number of metros and for unemployment rate further, the f-statistic showed the two variables to be jointly significant.

The results found in our model have some public policy implications. Having a public higher education program could prove to be beneficial to the state economy. The effects of education expenditures were consistently positive and significant, which would indicate that funding such a program is very likely to have positive returns. As to how extensive the program should be would be dependent on other factors, but if increasing income of the state is a main priority for the long-run, policy makers would do well to seriously considering increasing investments in post-secondary education.

Appendix

Figure 5:

. regress logavginc logeduexp

Source	SS	df	MS
Model	.112723986	1	.112723986
Residual	1.2015206	48	.025031679
Total	1.31424459	49	.026821318

Number of obs =	50
F(1, 48) =	4.50
Prob > F =	0.0390
R-squared =	0.0858
Adj R-squared =	0.0667
Root MSE =	.15821

logavginc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logeduexp	.1677659	.079057	2.12	0.039	.0088111	.3267207
_cons	9.343137	.6924944	13.49	0.000	7.950783	10.73549

Figure 6:

. regress logavginc logeduexp unemploy povrate post2nddegpercent logmetros

Source	SS	df	MS
Model	.901883041	5	.180376608
Residual	.412361548	44	.009371853
Total	1.31424459	49	.026821318

Number of obs =	50
F(5, 44) =	19.25
Prob > F =	0.0000
R-squared =	0.6862
Adj R-squared =	0.6506
Root MSE =	.09681

logavginc	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
logeduexp	.2096149	.0491527	4.26	0.000	.110554	.3086757
unemploy	.0240728	.0082977	2.90	0.006	.00735	.0407957
povrate	-.0257907	.0062283	-4.14	0.000	-.0383429	-.0132384
post2nddeg~t	.0096326	.0029162	3.30	0.002	.0037553	.0155099
logmetros	-.025386	.0197677	-1.28	0.206	-.0652252	.0144533
_cons	8.803708	.485613	18.13	0.000	7.825019	9.782397

Figure 7:

```
. regress logavginc logeduexp unemploy post2nddegpercent logmetros

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  4,    45) =   14.55
       Model |  .741182592     4  .185295648           Prob > F      =  0.0000
    Residual |  .573061997    45  .012734711           R-squared     =  0.5640
-------------+------------------------------           Adj R-squared =  0.5252
       Total |  1.31424459    49  .026821318           Root MSE      =  .11285

------------------------------------------------------------------------------
   logavginc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logeduexp |   .2103444   .0572963     3.67   0.001     .0949437    .3257452
    unemploy |   .0189234   .0095633     1.98   0.054     -.000338    .0381848
post2nddegpercent |   .0172957   .0026273     6.58   0.000     .0120041    .0225873
   logmetros |  -.0509381   .0218915    -2.33   0.025      -.09503   -.0068463
       _cons |   8.231848   .5426998    15.17   0.000     7.138794    9.324901
------------------------------------------------------------------------------
```

Figure 8:

```
. regress logavginc logeduexp povrate post2nddegpercent logmetros

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  4,    45) =   18.85
       Model |  .823003062     4  .205750765           Prob > F      =  0.0000
    Residual |  .491241527    45  .010916478           R-squared     =  0.6262
-------------+------------------------------           Adj R-squared =  0.5930
       Total |  1.31424459    49  .026821318           Root MSE      =  .10448

------------------------------------------------------------------------------
   logavginc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   logeduexp |   .2023014   .0529791     3.82   0.000      .095596    .3090067
     povrate |  -.0230827    .006646    -3.47   0.001    -.0364685   -.0096969
post2nddegpercent |    .008756   .0031305     2.80   0.008     .0024509    .0150611
   logmetros |  -.0047808     .01991    -0.24   0.811    -.0448817      .03532
       _cons |   9.029293   .5173433    17.45   0.000      7.98731    10.07128
------------------------------------------------------------------------------
```

Figure 9:

```
. regress logavginc logeduexp povrate post2nddegpercent

      Source |       SS       df       MS              Number of obs =      50
-------------+------------------------------           F(  3,    46) =   25.64
       Model |  .822373631     3  .274124544           Prob > F      =  0.0000
    Residual |  .491870958    46  .010692847           R-squared     =  0.6257
-------------+------------------------------           Adj R-squared =  0.6013
       Total |  1.31424459    49  .026821318           Root MSE      =  .10341

------------------------------------------------------------------------------
    logavginc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    logeduexp |   .2035674   .0521733     3.90   0.000     .0985479    .3085868
      povrate |  -.0237157    .006038    -3.93   0.000    -.0358695   -.0115619
post2nddegpercent | .0086236 .0030498   2.83   0.007     .0024847    .0147625
        _cons |   9.023339   .5114284    17.64   0.000     7.993887    10.05279
------------------------------------------------------------------------------
```

References

Breton, Theodore R. "The Role Of Education In Economic Growth: Theory, History And Current Returns." *Educational Research* 55.2 (2013): 121-138. *Academic Search Complete*. Web. 12 Mar. 2014.

Baldwin, J., Borrelli, S. A., & New, M. J. (2011). "State Educational Investments and Economic Growth in the United States: A Path Analysis." *Social Science Quarterly* (Wiley-Blackwell), 92(1), 226-245. doi:10.1111/j.1540-6237.2011.00765.x

Eid, Ashraf. "Higher Education R&D and Productivity Growth: an Empirical Study on High-Income OECD countries." *Education Economics*. 20.1 (2012): 53-68. Print.

"Household Income - Distribution by Income Level and State: 2009." *The Census Bureau*. n.p., n.d. Web. 10 Feb 2014.

"Metropolitan Areas and the Next Economy: State Profiles." *The Brookings Institution*. N.p., n.d. Web. 08 Apr. 2014.

"New State-by-State College Attainment Numbers Show Progress Toward 2020 Goal." *U.S. Department of Education*. N.p., n.d. Web. 07 Apr. 2014.

"Poverty." *Income, and Health Insurance in the United States: 2012*. N.p., n.d. Web. 08 Apr. 2014.

"Unemployment Rates for States." *U.S. Bureau of Labor Statistics*. U.S. Bureau of Labor Statistics, n.d. Web. 08 Apr. 2014.

"Resident Population Data (Text Version)." - *2010 Census*. N.p., n.d. Web. 08 Apr. 2014.

"State and Local Financial Support for Higher Education by State: 2009-2010." *The Census Bureau*. n.p., n.d. Web. 10 Feb 2014.