

US Unemployment Rate Associated With Economics

Posted on September 2, 2019

A. Here is my idea. Others had similar ideas, but here is my understanding of the theoretical basis of my model.

My idea is to consider state differences in personal income, not personal income per person, but total personal income earned in a state. I was interested in this because, in the Wall Street Journal, that personal income earned in a state depends on educational levels of attainment and other things like income distribution. In terms of economic theory, I read in my Econ 200 textbook that such might be true, and I found an example online that said that this might be true. After discussing my plan my plan with my professor, I began to work. We decided to use the total personal income earned in a state as the left-hand side variables, with educational attainment, income distribution, and employment growth variables on the right-hand side. The discussions with my professor are found in my emails, which are printed and attached to my assignments.

B. Here is my data.

Getting data was tough, but with some effort, I got most of the data items. These are found in the example EXCEL file named Project Expectations_EX1.XLSX. The hardest part is that some data is available for slightly different years and in different formats. Some choices had to be made. The biggest choice is what “order” of states to use as sometimes that data are listed by state name, by state abbreviation, and by state region. This took some time and was difficult. When I brought these difficulties to the attention of my professor, suggestions were made that proved helpful.

My main sources of data are the US Census and the Bureau of Labor Statistics. (U.S. Census Bureau, n.d.; U.S. Bureau of Labor Statistics, n.d.-a, n.d.-b).

The personal income data is available in a format organized by state region. (U.S. Bureau of Economic Analysis, n.d.). Details are in the attached EXCEL file under the tab “Personal Income.” Same for “Educational Attainment”, “Income Distribution” and “Employment by State”. I decided to use the order of states shown in the “Employment by State” tab. This took a lot of time and was painful.

From the Employment By State data set, I calculated the following (in the Employment By State tab).

- TotEmp is the total employment in the state in January 2017.
- EmpGrowth is the % change in employment from January 1976 to January 2017.'
- Urate is the employment rate in July 2016.
- From the Income Distribution data (in the Income Distribution tab), I calculated the following:
- Poor% is the percentage of the population with an income below \$17,499
- Rich% is the percentage of the population with an income above \$97,500

From the Personal Income Data, I got a variable called Persians, which is the total personal income earned in a state in 2016. This data is in the Personal Income tab.

For this data set, I show the average, minimum, and maximum for each variable below. This example is interesting, but if there were a real project topic, I might be interested in listing the top 5 and bottom five states. Also, the table formatting is not up to standard.

This calculation is in the "CompileDataResults" tab. See the figure below:

	Average	Minimum	Maximum
Persinc	311,782	31,317	2,172,056
TotEmp	2,978,114	287,642	18,175,968
EmpGrowth	81.72%	13.88%	370.66%
Urate	4.5	2.7	6.7
Poor%	3.34%	2.03%	4.20%
Rich%	6.46%	2.88%	15.66%

Here is my correlation matrix. It is calculated in the CompileDataResults tab. Notice I had to manually put the variable names in the table in EXCEL. I see very high correlations between "Persinc" and "TotEmp." At this point, I am not surprised because the more people working, the greater the personal income earned. I am a bit surprised by the positive correlation between Urate and Personc, are you?

	Persinc	TotEmp	EmpGrowth	Urate	Poor%	Rich%
--	---------	--------	-----------	-------	-------	-------

Persinc	1					
TotEmp	0.990686	1				
EmpGrowth	0.073623	0.10726769	1			
Urate	0.200919	0.22005478	0.018049	1		
Poor%	-0.03983	-0.0024651	-0.05773	0.220876	1	
Rich%	0.190634	0.13454664	-0.15527	-0.08234	-0.54593	1

C. Here is the model I will estimate

a. My left-hand side or dependent variable is Persinc the total personal income earned in a state.

b. My right-hand side dependent variables are things I expected to see related to Persinc in a way that reflects cause.

- Totemp is total employment in the state.
- EmpGrowth is growth in total employment over a long period; I would expect that personal income is higher in growing states than in non-growing states.
- Urate is the state unemployment rate. While this is defined for only one period I suspect that would see that as unemployment increases (as a % of labour force) total personal income would increase across states. The weakness of this variable is that it is for one period of time. Urate might also be a “control” variable, meaning that it sort of measures labor market conditions in the background.
- Poor percentage should be associated with lower personal income earned in a state. Sure, lower earners (on the income scale) make lower incomes. However, it is also true that poorer workers as a group may represent lower accumulations of [human capital](#), preventing employment in higher-value occupations.
- Rich%, I would expect that higher income earners add more to personal income, and there are direct and indirect reasons for this. Higher-income workers are concentrated in higher-earning industries, which would benefit the entire economy of the states in which they are located.

The model estimated has the following form

$$\text{Persinc} = B_0 + B_1 \cdot \text{TotEmpGrowth} + B_3 \cdot \text{Urate} + B_4 \cdot \text{Poor\%} + B_5 \cdot \text{Rich\%} + u$$

Where I denote state I (one of the states) and u is the random error term

all regression models.

D. These are my estimated results.

a. I estimated the model using the Data tool “regression” in EXCEL.

Regression Statistics	
Multiple R	0.992717
R Square	0

	Coefficient	T Stat
Intercept	-44740.6	-0.50411
TotEmp	0.113809	52.34238
EmpGrowth	-14507.4	-1.30951
Urate	-4293.85	-0.59675
Poor%	-293910	-0.15277
Rich%	905514.8	2.022896

The model is shown to be:

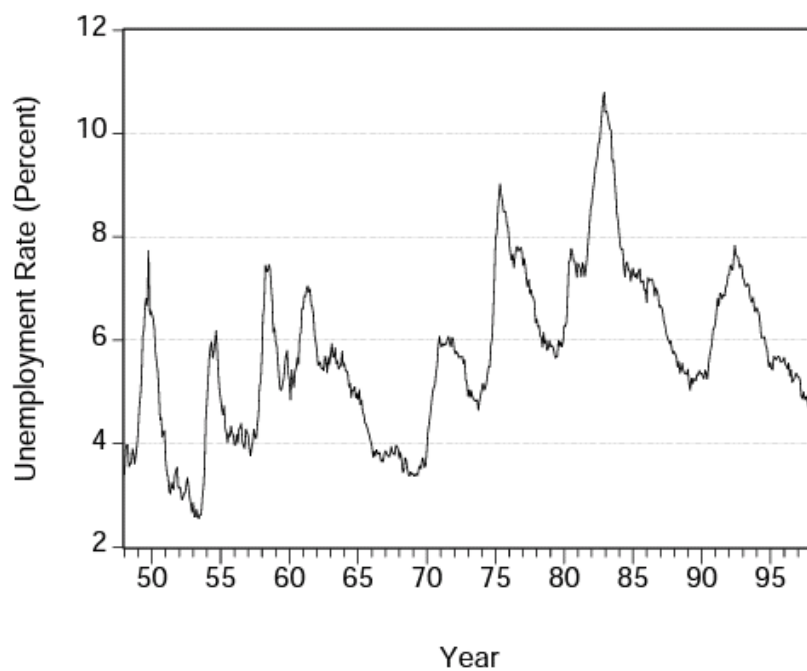
The overall fit is very high, and 98% of the variation across states in Persinc is explained by this

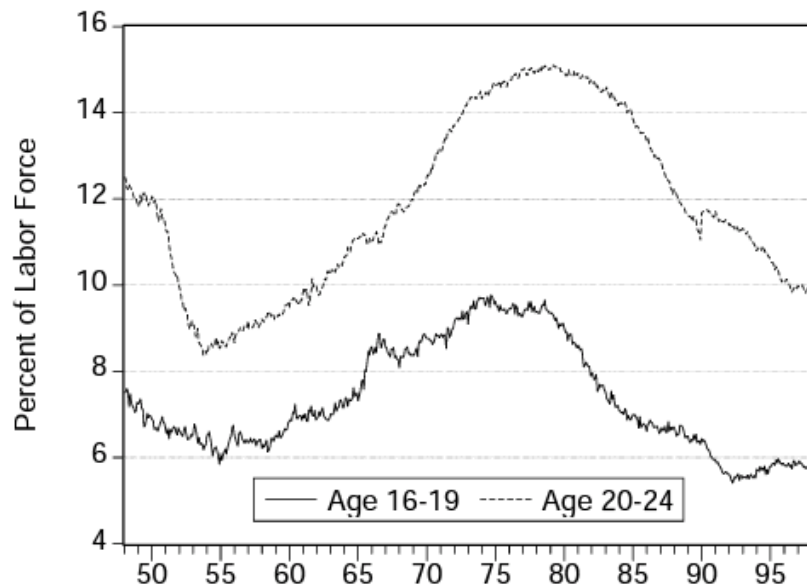
estimated model. The F statistic also shows a significant relationship.

The model is:

$$\text{Persinc} = -44740 + 0.0114 \cdot \text{TotEmp} + (-14,507.4) \cdot \text{EmpGrowth} + (-4,293.85) \cdot \text{Urate} + (-293,910) \cdot \text{Poor\%} + 905514.8 \cdot \text{Rich\%} + u$$

The graph below shows the residual values for each state.





The residuals are small, but in terms of trends, we see that the states with higher personal income have larger residuals. This indicates that our model is not exactly the best it could be. It likely suffers from Heteroscedasticity, an error issue that is beyond our scope at this point. I calculated the residuals by creating the fitted values in the spreadsheet using a standard formula approach. I then plotted them in a graph in the Residual Graph tab.

Describe the individual estimated parameters:

The individual coefficients seem different from expectations. I expected a positive sign for TotEmp and got one. There is no reason to expect EmGrowth to have a negative sign, but that is the result. I expected Urate to have a negative sign, and it does, and my expectations are met for Poor% and Rich%.

Only two of the variables are statistically different from zero: TotEmp and Rich%. This is an odd result in that the overall fit is very good, but few coefficients are statistically different from zero.

In summary, I thought up and estimated a model. I had a guide as to what to expect because of Economic Theory and work done by others. The data gathering was tough to do, and I am not sure that the data is accurate in that it might cover different periods. For example, I might have used a more recent period of employment rate.

My estimated model had a great fit, well over 94% for the R-squared. Some of the coefficients had the

expected or reasonable sign.

After showing these results to my professor, I saw that part of the estimation problem of insignificant variables might be due to the high correlation between TotEmp and Persians. Further, I don't have a great reason to include TotEmp in the equation. These results are in the "Alternative Results" tab of the spreadsheets. This did not improve the results; the R=squared is much lower, and again, a few coefficients are statistically significantly different from zero.

Work Cited

The main source is the US Census and the Bureau of Labor Statistics.

References

U.S. Bureau of Economic Analysis (n.d.). *Personal income by state*

U.S. Bureau of Labor Statistics (n.d.-a). *How the government measures unemployment*

U.S. Bureau of Labor Statistics (n.d.-b). *Local Area Unemployment Statistics*

U.S. Census Bureau (n.d.). *Educational attainment*