SSC442_Final

Riley Latham

Overview

This paper attempts to predict unemployment results from a number of sources and analyze the inputs which effect it most. The goal is to find inputs we're able to manipulate in order to reduce unemployment in previously unexplored and novel ways. We use data from the United States census as well as the Beaureu of Labor Statistics to analyze inputs such as air quality, minimum wage, union density, and more. Previously many thought raising minimum wage was a credible way of lowering unemployment, our analysis finds that while it does have an effect on poverty it is not a credible tool for combatting unemployment. Alongside this we find minimal effects on unemployment from CO2 emissions and air quality improvements, SNAP benefit recipients, and manufacturing employment. We originally suspected these may play an important role in unemployment. Our analysis highlights inputs that are most effective in combatting unemployment and includes policy suggestions as well.

1. Introduction

Unemployment is a rising concern due to the recent outbreak of the coronavirus as nearly 26 million Americans have applied for unemployment assisstance. However, we have good data on unemployment by state and are able to analyze novel ways of assisting the workforce. Many of the considerations are based in economic theory which gives validity to the empirical work of our analysis. The work conducted helps highlight options for the government to pursue that can impact unemployment in a meaningful way that has potentially been overlooked prior.

We use panel data from the Beaureu of Labor Statistics and the United States census to perform a time series analysis and estimate what inputs greatly impact unemployment. Most of the inputs we're interested in are obvious, but we examine novel inputs as well such as air quality by proxy and employment data by sector. The data allows us to look through both monetary and non-pecuniary factors ranging from income to CO2 emissions and construct regressions that ultimately guide future policy. We use data from the year 2000-2019 as it contains the variables we are most interested in with relative ease of access. This data is publically available and can be found at _____WEBSITE_____.

We find somewhat disappointing results from a naive analysis, despite a very high $R^2 = 0.96$. Nearly all inputs are statistically insignificant, which leads us to believe that we have very high collinearity among our independent variables. Collinearity can be from including unnecessary explanatory variables in our regression that are related to one another. This could be something like income by household and income per capita. Due to these findings we use a step function to find stronger regressors while keeping a high R^2 value. After performing a regression with our limited inputs we find the much slimmed down regression has statistical significance in the poverty rates, per capita income, and CO2 emissions. All of these are explored in the results and discussion sections.

Above is a brief overview of the project which is explored in depth below. Section 2 covers the data collection, feature selection, and methods used for analysis. Section 3 reviews the results from the analysis and examines the most important findings. Section 4 is a discussion about the usefulness of the findings and possible future policy that could be developed from the results.

2. Data and Methodology

The following section outlines the data collection process as well as the methods used for analysis. We discuss the uses and reasoning behind our feature selection and the interpretation of our outputs.

2.1 Data

We collected data from the 2000-2019 United States census as well as the Beaureu of Labor Statistics which are both publically available datasets that we've merged by state to perform a time-series analysis. We've decided to grab the following input variables for our analysis using primarily economic theory as a guiding tool as well as any interesting variables we felt might have an under-appreciated effect on unemployment. The inputs are Household Income, Per Capita Income, Minimum Wage, Housing Price Index, Manufacturing Employment, Labor Force Participation, GDP, SNAP Benefit Recipients, Poverty Rates, and CO2 Emissions.

Unfortunately the datasets represent high level statistics and we are unable to access more granular data such as unemployment by county. This is a natural limitation of our data sources, but is not as limiting as it may first seem. Most unemployment policies are done at a state or even federal level which are then controlled for by year. Due to the implementation and scope of previous unemployment policy it is best to analyze data with a broad reach to provide more accurate insight into the effects it will have.

2.2 Methods

The first regressions we perform are a naive regression of unemployment on all of our data. This preliminary search is performed to determine the data's ability to predict unemployment before further analysis. We find that the data is able to predict unemployment very well, but has high collinearity issues. This occurs when indepedent variables are strongly correlated with one another and presents us with a problem. We must find a way to drop variables in a systematic way so that we keep the underlying predictions strong while letting us interpret the coefficients on our inputs more accurately. Once we're able to do this we can find the best possible inputs and use our regressions to make policy suggestions.

To solve the collinearity problem we use a step function that runs multiple regressions in order to minimize the relative AIC value of the regressions. The step function takes in a set of data and uses adding and dropping of independent variables to select the strongest model. AIC is a criteria used to measure the strength of the regression relative to the other regressions that could be run using the same data. When this is minimized it generates the strongest regression results with the least amount of input required. The step function has a limitation in that it can only output linear models that do not have polynomial terms or interaction terms. This should be noted and cared for in all uses of the step function.

3. Results

Below we examine the results from our naive regression before using the step function to get statistically significant results. These results are used to guide our discussion in section 4.

3.1 Naive Regression

The naive regression is shown in table 1 under the appendix.

The naive regression is performed to see if the dataset provides an accurate prediction for unemployment. We can see a very high R^2 value of 0.96 leading us to believe the data is reliable. Unfortunately it has very low statistical significance across nearly all of our variables so our analysis would likely produce very little meaning, this finding is supported by the low F-statistic for our model. Despite explaining much of the variation we have little significance to the model. This problem can often stem from the issue of collinearity, when two our more of our independent variables are very similar. Collinearity introduces bias into our regression which renders any policy suggestion from the analysis shakey at best. Luckily this is a common problem with many possible solutions. We solve this problem below using a step function.

3.2 Step Function

After running our step function we see a much slimmed down model providing us with statistically significant findings. The variables the step function keeps in our regression are Year and State for controls, Employment, Household Income, Per Capita Income, Union Membership Percentage, Overall Homeless, Houseing Price Index, GDP, and SNAP Benefits. We suspect it keeps both income measures as a pseudo polynomial term. They come from different data sources which avoids perfect collinearity issues and the step function is unable to output regressions with non-linear effects. Interestingly it drops out variables that we thought would have much more importance such as minimum wage and the CO2 emissions. This tells us they must have a small impact on the unemployment rate, although this could be due to lack of supporting controls.

In the model we interpret in our discussion section we remove the Household Income term and instead use a 2nd degree polynomial interpretation of Per Capita Income. This leaves us with a very similar regression but with much more easily interpreted coefficients that are still statistically significant.

Both models are in the table 2 in the appendix.

4. Discussion

This section provides space for a discussion of our results broken into policy suggestions and possible improvements to the data and methods of analysis. We consider the key results shown above and provide policies that could reduce unemployment.

4.1 Policy Suggestions

4.1.1 Minimum Wage

Now that we've performed our regression analysis we can begin crafting policy suggestions to help reduce unemployment in novel ways. First we notice that Minimum wage is no longer in our regression because of the statistical insignificance. While many politicians and workers rights groups cite raising the minimum wage as a primary campaign issue we find little relevance as to the impact on unemplyment. Perhaps it is meant as a tool for reducing poverty, we find a 1 percent increase in the minimum wage can increase per capita income by 4.65%, however, this doesn't translate to unemployment directly. Minimum wage has no statistically significant effect on unemployment, but can impact unemployment through its effect on per capita income. This may lead us to believe that we can use minimum wage as a tool to reduce unemployment as well. We see in table 2 that a 1% increase in per capita income decreases unemployment by 0.26%, so a 1% in the minimum wage will lower unemployment by an expected 1.18%. This would lead us to believe that minimum wage is not a strong economic tool and is perhaps more often used for political reasons than its effect on the economy. It's unfortunate that we don't have an easy way of influencing unemployment, but we're able to find more novel ways to make a change.

4.1.2 SNAP Benefits

One interesting correlation we find is between SNAP Benefits and unemployment. We believe this is actually because SNAP benefits are simply positively correlated to the unemployment rate and not that they somehow cause unemployment, but we explore the possibility regardless. Consider a city with stronger SNAP style benefits that allow for the unemployed to have a better diet. This safety net method lessens the impact of becoming unemployed in this city meaning people are more open to taking risks. In turn this could cause people to take larger loans on housing, which also has a significant impact on unemployment. With enough people making risker financial decisions because of the relatively lower risk due to SNAP benefits being readily given to the unemployed those who fail to pay back their loans may end up unemployed. This compounding effect of taking on risker financial decisions could have an impact on the overall unemployment rate. If this is the case it would be wise to consider cutting back on SNAP benefit spending or becoming more stringent with the requirements for getting SNAP.

4.1.3 CO2 Emissions

Another interesting correlation is between CO_2 emissions and unemployment. We believe this is actually because cities with lower CO_2 emissions are more liberal and have better unemployment outreach, but we explore this correlation as a way of reducing unemployment in a novel way. Consider a city with less air pollution, in general the people living there will be more happy with the quality of life. In turn this drives up housing prices as more people choose to live in this place, but this helps pull business to the growing community. As a distant effect we see that cities with less pollution draw in more people and they must have jobs there as well. In urban economics this is known as agglomeration, the slow compounding effects of growing cities. It's very difficult for us to establish what the new unemployment rate would be because of the simultaneity that comes with agglomeration effects, but our regression helps shine some light on the possible effects.

From above we can see that a 1% decrease in CO_2 emissions will decrease unemployment by 0.04%. This is an incredibly small effect, but as stated before the agglomeration effects could increase this effect qutie a bit. It's because of this we believe exploring methods to improve the environment could decrease unemployment. This has the added benefit of directly increasing employment through the city to implement these environmental improvements.

4.1.4 Per Capita Income

We see from our regression that per capita income has the largest effect on unemployment. This makes sense in many ways because people with more money end up spending a portion of the additional income which helps bolster the local economy. We can see that an increase in per capita income leads us to massive decreases in unemployment and so we believe this is the best strategy for reducing unemployment and explore different implementation stratagies here.

One possible method of implementation is to subsidize employment through local business. This makes it easier to hire more employees and grow their business which helps directly lower unemployment. Not only does it have this immediate impact it also helps to put money into the workers pockets which they then put back into the economy. This cycle is similar to the agglomeration we discussed above in that it has a compounding effect and ultimately leads to a new equilibrium in the employment market. This is perhaps the most easily implemented strategy from a logistics standpoint as well. Targetting and subsidizing business is something that is done at every level of government and is well understood.

Another more recently debated strategy is a universal basic income. This was popularized by scandanavian countries and seen in America through the oil checks given to Alaskans; more recently it was a campaign promise from presidential runner Andrew Yang. This promising system is able to effect every person, both employed and unemployed, and because of this it's a great strategy for improving the overall economy. This does have the downside of not having a direct impact on unemployment, but depending on the spending habits of the population this could be a much more powerful effect. By stimulating the economy it creates new space for business to expand which provides employment opportunities. This strategy would do best if it is used along with something like a government jobs program.

4.2 Improving Results and Limitations

The primary limitation for this work came from the scarcity of data dating back before 2005. In all states sampled we find varying levels of data tracking for the union variables as well as SNAP. This could be due to expansion of these programs or because the laws surrounding them have walled off data to the public. Because of these often one or two datapoints being omitted we lose the entire year/state of information, which could have an impact on the regressions outcomes. We still have a decently sized dataset and so we keep all the necessary econometric assumptions we need for making strong regressions.

Appendix

<div align="center">

Table 1:

	Dependent variable:
	Unemployment.Year.Average
Year	0.428
	(0.275)
StateIndiana	−5.034
	(5.548)
StateOhio	31.240***
	(6.766)
Employment	−0.170**
	(0.061)
Household_Income	0.069
	(0.051)
Per_Capita_Income	−0.268***
	(0.067)
Union.membership.percentage	0.567
	(0.772)
Union.Coverage	−0.219
	(0.658)
Overall.Homeless	−0.029
	(0.025)
Min.Wage	−0.400
	(0.762)
House.Price.Index.Aggregate	0.068**
	(0.030)
Manufacturing.employmet	−0.019
	(0.055)
Labor.Force.Participation	−0.266
	(0.322)
GDP	−0.647***
	(0.113)
SNAP_Benefits	0.191**
	(0.087)
Poverty_Rates	−0.020
	(0.491)
CO2_Emissions	0.0004
	(0.021)
Constant	−834.832
	(556.721)
R^2	0.962
Adjusted R^2	0.932
F Statistic	31.548*** (df = 17; 21)

*p<0.1; **p<0.05; ***p<0.01

</div>

Table 2:

	Dependent variable:	
	Unemployment.Year.Average	
	(1)	(2)
Year	0.068	0.320**
	(0.206)	(0.150)
StateIndiana	−2.832	7.665**
	(3.583)	(3.323)
StateOhio	36.174***	33.881***
	(4.836)	(6.182)
Employment		−0.117*
		(0.066)
Household_Income		−0.008
		(0.058)
log(Per_Capita_Income)		−0.886*
		(0.457)
GDP	−0.739***	−0.715***
	(0.107)	(0.135)
poly(log(Per_Capita_Income), 2)1	−29.571***	
	(5.480)	
poly(log(Per_Capita_Income), 2)2	−9.377***	
	(2.366)	
House.Price.Index.Aggregate	0.091***	0.038
	(0.022)	(0.029)
SNAP_Benefits	0.342***	0.158
	(0.063)	(0.095)
Union.membership.percentage	0.302	0.336
	(0.235)	(0.264)
Overall.Homeless	−0.063**	−0.045
	(0.024)	(0.032)
CO2_Emissions	−0.040*	
	(0.023)	
Constant	−144.273	−637.468**
	(415.729)	(298.339)
R^2	0.934	0.907
Adjusted R^2	0.908	0.869
F Statistic (df = 11; 27)	34.972***	23.935***

*p<0.1; **p<0.05; ***p<0.01