Data Project 2

ECO220 - Introduction to Data Analysis and Applied Econometrics

Dr. Nick Zammit

April 18, 2022

Submitted by:

Gurleen Kaur - 1007317976

Moaaz Ali - 1007438774

Mohammed Yusuf - 1006695783

Muhammad Farooq - 1006331275

Introduction

Our data project focuses on the data extracted from 2009 to 2020 through STATA to work on a model that gave us information post 2008-2009 Financial crisis. We figured it would be an interesting analysis to work with the numbers after the financial crisis as it would allow us to learn more about the crisis through this data project and we were hoping to find an engaging analysis. The variables used in our regression model includes minimum wage, annual GDP, quarter GDP, emigrants, immigrants, population, CPI, labour force which are the independent variables compared to the dependent variable of unemployment15over and their effects around unemployment. For the model the provinces taken into account are Ontario, Quebec and British Columbia (based on the population size respectively). The provinces chosen allow us to streamline our data and work with numbers that represent the majority of the Canadian population. Based on the numbers provided in the data our group wanted to come up with a model that chooses a large segment of the Canadian population and the years that cover the most accurate and recent information from the data. Taking into consideration all the factors our conclusion model covers the three provinces and the niche time period of post financial crisis 2008-2020. The model is unique and unconventional as it captures the regression for a niche time period for a vast majority of the population.

Analysis

The question being assessed is that does minimum wages impact unemployment ? Does the relationship differ between region, group or time? The relationship between unemployment and minimum wage in our data project is being analysed during the period of financial crisis (2009-2020). For the three provinces we generated both simple and multiple regression.

Simple Regression

For our simple regression the variables we took into account were unemployment15over and minimum wage. Unemployment15over being a dependent variable and minimum wage being the independent variable a simple regression equation 1 being formed from the extract of table 1:

$$\text{Unemployment15andover} = B_0 + B_1 \text{ (minimum wage)}$$

Where B_0 represents the y intercept and B_1 denotes the unit change between minimum wage and unemployment15over. While we regressed unemployment15andover and minimum wage in STATA our result demonstrated that there's a negative correlation between the two. Another point of significance to be noted is that the $P > | t |$ is equal to 0.
As simple regression compares unemployment15andover and minimum wage, these 2 variables being compared alone are not sufficient enough to display a concrete or accurate model. Other than just the minimum wage there are other several variables that need to be included in the model to get a better picture of the economy as a whole.

Multiple Regression

While we add the other variables in the model it converts the simple regression into multiple regression. The equation 2 being formed in the multiple regression is as follows:

$$\text{Unemployment15andover} = B0 + B_1 (\text{minimumwage}) + B_2 (\text{Dummy Variable 2}) + B_3 (\text{Dummy Variable 3}) + B_4 (\text{Dummy Variable 4})$$

In here B_2, B_3, B_4 represents the unit change between unemployment15andover and the dummy variables. Dummy variables in the equation are the replacements of the best fit variables on which our model will be built.

Trials

We tried and tested each variable given to us in the data to come up with the best fit model . A perfect model would have an adjusted R^2 with a high value, p-value of near to zero and the three tests that model needs to pass are ovtest, hettest and vif test. Our journey of finding the suitable model began by
Trail 1: choosing population as dummy variable number 2. Our sole reason for taking population as dummy variable number 2 was because unemployment is more or less affected by the change in the population. Replacing dummy variable with population equation #3 is as follows.

$$\text{Unemployment15andover} = B0 + B_1 (\text{minimumwage}) + B_2 (\text{population})$$

With population being taken as dummy variable 2, as shown in figure (2) it did not pass the ovtest.

Trail 2: Annual GDP represents the total amount of spending in a country. There is a negative relationship between annual GDP and unemployment, which fits our model with the minimum wage so our trial 2 included annual GDP as dummy variable 2. Equation #4 with annual GDP

$$\text{Unemployment15andover} = B0 + B_1 (\text{minimumwage}) + B_2 (\text{annual GDP})$$

The model passed the criterion of adjusted R2 at a value of 0.7378, p value being at 0, a negative correlation and passing the ovtest and hettest.

Our new equation will look like:

$$\text{Unemployment15andover} = B0 + B_1 (\text{minimumwage}) + B_2 (\text{annual GDP}) + B_3 (\text{Dummy Variable 3}) + B_4 (\text{Dummy Variable 4})$$

Trial 3: In order to find the best fit for dummy variables 3 and 4 we had an extensive trial where we used new variables like

- GDP per capita: - as annual GDP was already taken into consideration the GDP per capita did not work as a suitable dummy variable 3. We rejected the condition of including it in our model.

- Emigrants: - Emigrants were not taken as dummy variable 3 as it caused a conflict in the p value as shown in table (. , .) so they were rejected in the model.

- Labour force, CPI, growth rate, GDP per capita: regressing these as dummy variables 3,4 and 5 our model could not pass the p value test, because the p value being significantly high at 0.620 got rejected to pertain in our model.

Trial 4: **The Best Model**

After regressing many trials and including a lot of variables with several combinations in dummy variables our group came up with the best model which passed the ovtest, the hettest and the VIF test. The equation for the best fit model consisted of the following variables.

$$\text{Unemployment15andover} = B0 + B_1 (\text{minimumwage}) + B_2 (\text{annual gdp}) + B_3 (CPI_{2002100}) + B_4 (\text{immigrants})$$

From the extract of the table (...) our adjusted R^2 was of a higher value consisting of a value of 0.7966, p value being the least, all pertaining to the negative relationship with unemployment15over. The null hypothesis of the model having no omitted variables (ov test) is also passed. The heteroskedasticity with Chi^2 with a value of 0.21 less than the p value also passes the test. To pass the multicollinearity test, the mean VIF has to be less than 10 ($7.8 < 10$). Therefore all the conditions to pass a strong regression model have been met.

Conclusion

In a nutshell the simple regression model did not include all the parameters required for a strong model. By adding more variables we proceeded with multiple regressions to get a clear understanding of the model analysed in our data project.

To build a successful model, we set a criteria to pass the ov test, the heteroskedasticity test, VIF test and also have a p value closer to zero and higher adjusted R^2 value. The model we built tested and explained the analysis by many forms of trial and testing on STATA. After extensive testing and manipulation with the variables we came across a model that fits perfectly well with the expectations we had set when starting the project.

Tables and Figures

```
. //Simple Regression
. regress unemploymentrate15over minimumwage
```

Source	SS	df	MS	Number of obs	=	132
				F(1, 130)	=	173.69
Model	107.181729	1	107.181729	Prob > F	=	0.0000
Residual	80.2203919	130	.617079938	R-squared	=	0.5719
				Adj R-squared	=	0.5686
Total	187.402121	131	1.43055054	Root MSE	=	.78554

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.6136127	.0465591	-13.18	0.000	-.7057243	-.5215011
_cons	13.43027	.4987119	26.93	0.000	12.44363	14.41691

Table 1

```
. regress unemploymentrate15over minimumwage population // Remove Population due to OVTEST
```

Source	SS	df	MS	Number of obs	=	132
				F(2, 129)	=	245.91
Model	148.461539	2	74.2307693	Prob > F	=	0.0000
Residual	38.9405825	129	.301864981	R-squared	=	0.7922
				Adj R-squared	=	0.7890
Total	187.402121	131	1.43055054	Root MSE	=	.54942

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.7394449	.034296	-21.56	0.000	-.8073003	-.6715895
population	1.59e-07	1.36e-08	11.69	0.000	1.32e-07	1.86e-07
_cons	13.3571	.3488633	38.29	0.000	12.66686	14.04733

Table 2

```
. //Omited variable//
. *Issue: omit a variable that is correlated with both X and Y --> violate the assumption of X and e being inde
> pendent --> bias
. // TEST
. estat ovtest // Fail

Ramsey RESET test using powers of the fitted values of unemploymentrate15over
       Ho:  model has no omitted variables
                 F(3, 126) =      3.59
                  Prob > F =      0.0156

. estat hettest // pass
```

Table 3

```
. regress unemploymentrate15over minimumwage annualgdp
```

Source	SS	df	MS	Number of obs	=	132
				F(2, 129)	=	185.27
Model	139.007659	2	69.5038295	Prob > F	=	0.0000
Residual	48.3944622	129	.37515087	R-squared	=	0.7418
				Adj R-squared	=	0.7378
Total	187.402121	131	1.43055054	Root MSE	=	.6125

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.7890272	.0409949	-19.25	0.000	-.8701366	-.7079178
annualgdp	2.82e-06	3.06e-07	9.21	0.000	2.21e-06	3.42e-06
_cons	14.1131	.395854	35.65	0.000	13.32989	14.89631

Table 4

```
. regress unemploymentrate15over minimumwage annualgdp emmigrants
```

Source	SS	df	MS	Number of obs	=	132
				F(3, 128)	=	122.90
Model	139.109592	3	46.3698639	Prob > F	=	0.0000
Residual	48.2925294	128	.377285386	R-squared	=	0.7423
				Adj R-squared	=	0.7363
Total	187.402121	131	1.43055054	Root MSE	=	.61424

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.7975031	.0442272	-18.03	0.000	-.8850142	-.7099919
annualgdp	3.01e-06	4.88e-07	6.18	0.000	2.05e-06	3.98e-06
emmigrants	-.0000204	.0000392	-0.52	0.604	-.000098	.0000572
_cons	14.20191	.432189	32.86	0.000	13.34676	15.05707

Table 5

```
.
. // Trying to add new variable Annual GDP 3
. gen growthrate=(annualgdp-annualgdp[_n-1])/(annualgdp[_n-1])
(1 missing value generated)

. regress unemploymentrate15over minimumwage cpi2002100 labourforce growthrate gdppercapita // not immigrants a
> nd annual gdp
```

Source	SS	df	MS	Number of obs	=	131
				F(5, 125)	=	269.63
Model	171.500231	5	34.3000462	Prob > F	=	0.0000
Residual	15.9017029	125	.127213623	R-squared	=	0.9151
				Adj R-squared	=	0.9118
Total	187.401934	130	1.44155334	Root MSE	=	.35667

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.15445	.0583276	-2.65	0.009	-.2698876	-.0390125
cpi2002100	-.0727461	.0134755	-5.40	0.000	-.0994157	-.0460764
labourforce	9.62e-07	5.24e-08	18.36	0.000	8.59e-07	1.07e-06
growthrate	-.152822	.3071787	-0.50	0.620	-.7607668	.4551227
gdppercapita	-111.342	11.43365	-9.74	0.000	-133.9706	-88.71334
_cons	21.20424	1.057884	20.04	0.000	19.11056	23.29792

Table 6

```
. regress unemploymentrate15over minimumwage cpi2002100 annualgdp immigrants //Best model 2 yet //immigrants
> migrants annualgdp   // Best Model yet
```

Source	SS	df	MS	Number of obs	=	132
				F(4, 127)	=	129.27
Model	150.451125	4	37.6127813	Prob > F	=	0.0000
Residual	36.9509962	127	.290952726	R-squared	=	0.8028
				Adj R-squared	=	0.7966
Total	187.402121	131	1.43055054	Root MSE	=	.5394

unemployme~r	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
minimumwage	-.3978967	.0890215	-4.47	0.000	-.5740542	-.2217391
cpi2002100	-.1016715	.0219878	-4.62	0.000	-.1451813	-.0581617
annualgdp	2.25e-06	6.73e-07	3.34	0.001	9.13e-07	3.58e-06
immigrants	.0000268	.0000134	2.00	0.047	3.01e-07	.0000534
_cons	22.24723	1.822102	12.21	0.000	18.64162	25.85284

Table 7

```
. estat ovtest

Ramsey RESET test using powers of the fitted values of unemploymentrate15over
       Ho:  model has no omitted variables
                 F(3, 124) =      1.83
                  Prob > F =      0.1453

. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of unemploymentrate15over

         chi2(1)      =     0.21
         Prob > chi2  =   0.6459
```

Table 8

```
. //Multicolinearity//
. *Issue: independent variables highly correlated with each other --> inflate the SE (harder to find significan
> ce)
. vif
```

Variable	VIF	1/VIF
cpi2002100	9.36	0.106798
annualgdp	7.97	0.125446
minimumwage	7.75	0.128973
immigrants	6.45	0.155117
Mean VIF	7.88	

Table 9