Using Econometric Methods to Analyze Javier Milei's Role in Argentina

Term Project for ECON415

Authored by Rowan Moxley

I. Motivation.

Javier Milei, the newest president of Argentina, stands as a clear shift in the nation's approach towards the macroeconomy; most notably, a shift from state neoliberal governance to near-radical libertarianism, complete with nationwide extreme government budget cuts, a promise to privatize education and healthcare, and the lofty goal of dollarization (with currency devaluation in rapid subsequence, in line with any economist's intuition.) In fact, Milei is on record an *anarcho-capitalist*; a political ideology which, although currently finding itself largely restricted to online chat servers, nevertheless did find its emergence into the mainstream with legitimate economic background such as Ludwig von Mises and P. Samuel Konkin III of the Austrian School. Konkin did, admittedly, go on to synergize the ideology known online as *agorism*, but agorist dogma aligns with that of anarcho-capitalism enough to where his contributions to the literature is quoted by those who aline with anarcho-capitalist ideals.

The goal of this paper is to use econometric methods, most importantly multiple ordinary least squares regression, to understand if Milei's economic policy can be proven to be beneficial to the Argentine economy. All statistical analysis will be conducted on provincial macroeconomic indicators in Argentina. If econometric procedure can prove that, say, government spending and education costs were the most significant predictors of real output, we can begin to understand if there is enough evidence to conclude that Milei's policies of reducing spending in these areas will be optimal for real growth in output for Argentines.

II. Core Assumptions.

The most important assumption of the model is that output (read: real gross domestic product) is the primary indicator for macroeconomic health. Although arguments exist for using inflation or unemployment in lieu of output, I maintain that output remains dominant as a predictor of the macroeconomy. Our understanding of the *Solow-Romer Model* (**Fig. 1.1**) of long-run economic growth in macroeconomic theory disqualifies price level and inflation as valid predictors, as they reflect only nominal changes and are irrelevant when compared to issues taken as real, i.e. investment, capital depreciation, total factor productivity, idea growth, etc. Similarly, the state-of-the-art *IS-LM-PC Model* (**Fig. 1.2**) of short-run economic growth uses price level and output as the main endogenous variables in question; inflation and interest rates (as a function of the money supply) are taken as functions of output. We can see that output is therefore just as good, if not better, as an indicator for short-run macroeconomic well-being due to its heralded status as the central endogenous variable within macroeconomic theory and linear correlation with other key issues of short-run macroeconomic concern.

Keen readers will have caught the central presupposition posited by my thesis, that output is most largely determined by activities within the country and not by a variety of international financial instruments and processes. This will be assumed for the purposes of this model, again, as the data I work with looks only at provincial data *within* Argentina. This is not without legitimate theoretical economic intuition, though. The *Rybczynski Theorem* (**Fig. 1.3**) within international economics posits that the natural endowments of any factor in production within a country determines the ratio of productive activities which use that factor intensively in production. As an intuitive extension, the standard *Heckscher-Ohlin Model* (**Fig. 1.4**) states that

countries will produce and export only those goods which they have a natural endowment in their factors-in-production. Understanding that only the 'export' component of aggregate consumption is utilized in real output figures, we can piece together that according to the state-of-the-art theoretical framework of international economics, a nation's GDP is mostly dependent on productive activities within their borders. This is useful for justifying output as a primary indicator of the macroeconomy, as it is something which Milei's domestic economic policy directly influences.

Figure. 1.1. Solow-Romer Model.

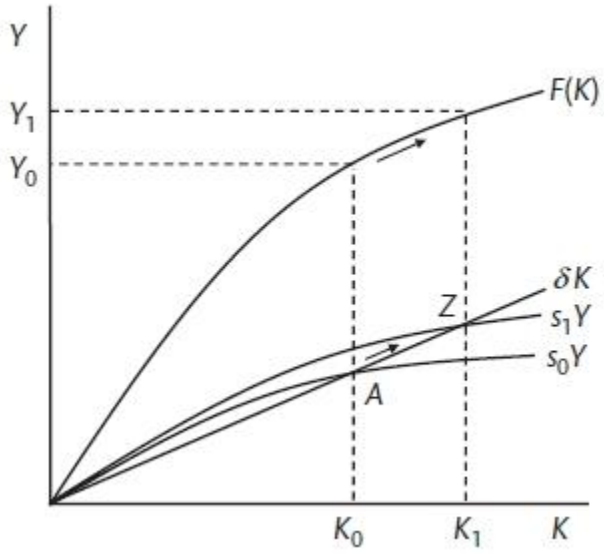


Figure. 1.2. IS-LM Model.

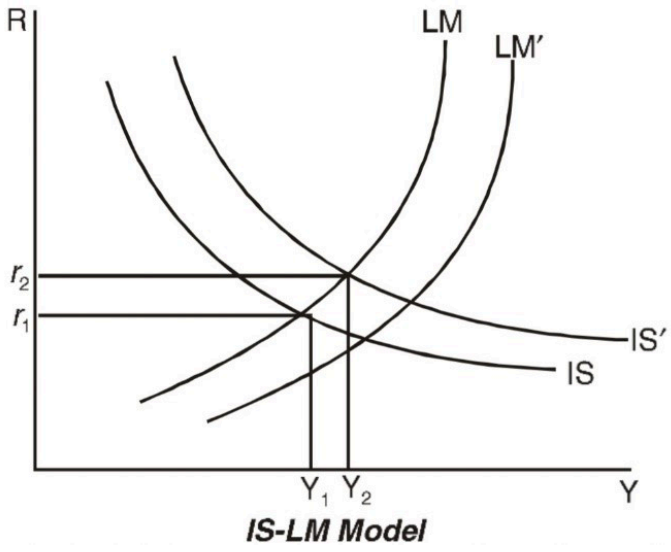


Figure 1.3. Rybczynski Theorem.

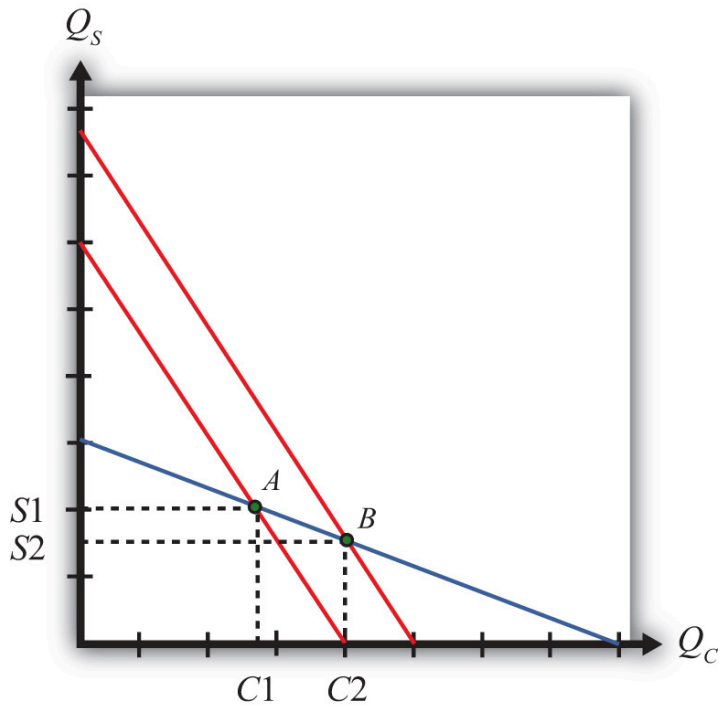
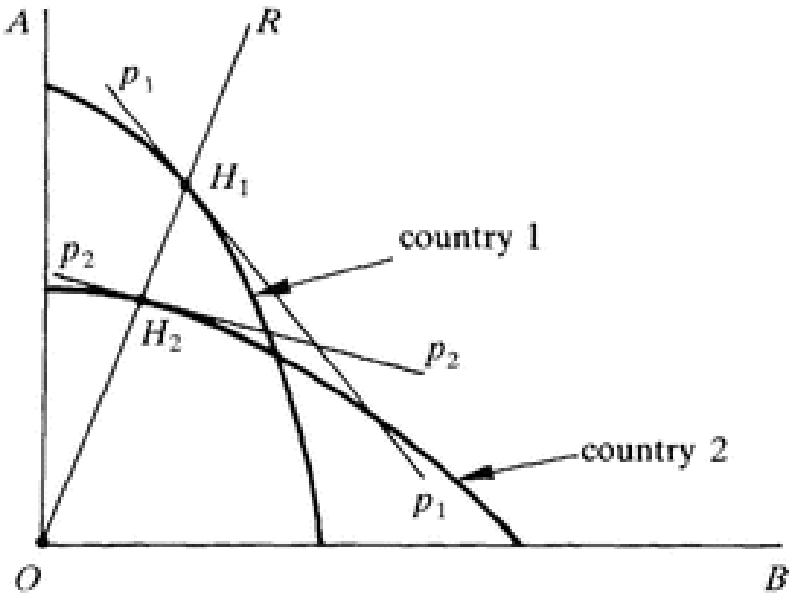


Figure 1.4. Heckscher-Ohlin Model.



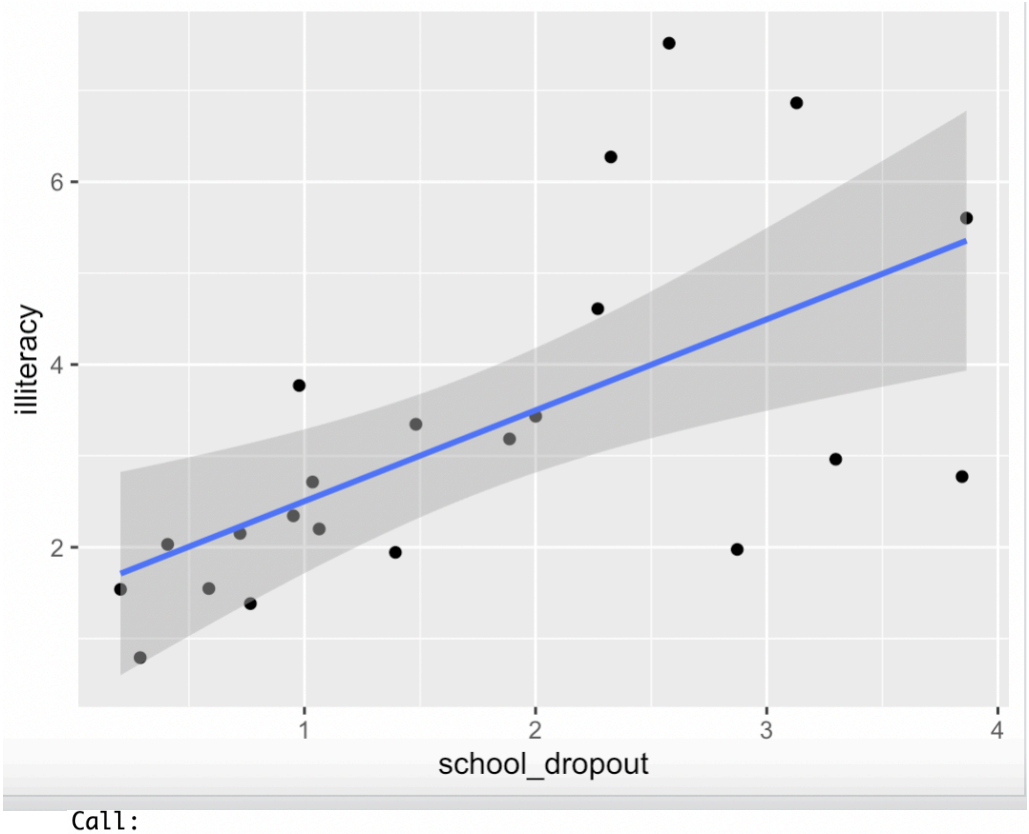
III. Data.

The data used for this project is "Argentina provincial data" posted on Kaggle by *Abid al Awan*. The initial data chopping involved a logarithmic transformation of the *gdp*, *population*, and *doctors_per_capita* variables. The idea is to express ideas like change in output, population, and doctors per capita as percentages, as this is much more interpretable and intuitive than raw single-unit increases. Specifically, there is a change of ($\Delta\widehat{y} = e^{\widehat{B1}} - 1$) percent from a one-unit change in a linear explanatory variable. Aside from the variables described above, all other variables of interest to the model are expressed as percentages. So, to follow a typical log-linear model of interpretation for one-unit increases in continuous variables, expressed as percentages, results in an elasticity graph with similar characteristics in interpretation as a log-log model. By log-transforming the non-percentage linear continuous variables, the model is set up to express every change as a percent change, greatly improving the interpretability of a model dealing with extremely large values for observations.

Two categorical variables were created to check Buenos Aires as an outlier, as economic intuition supports that a city that massive would skew the data in an unwanted way. The first categorical variable used *mutate(x = ifelse())* to check if any provinces constituted more than 40% of Argentina's GDP. Buenos Aires was the only province with the *1* reporting. This led to the birth of the *buenos_control* dummy variable; in effect, it creates a reference group with only Buenos Aires, as the results of interest are predictors for all of Argentina, not just Buenos Aires. This follows intuition as Buenos Aires is far and away an international city, subject to an entirely different set of forces that is not shared between the other 21 provinces. So, any regression performed with the *buenos_control* variable will control for Buenos Aires.

The next step involved selecting the key variables to study. The initial presumption was to select variables that could be affected by Milei's economic approach, the first three being *school_dropout, poverty,* and *illiteracy*. Illiteracy and school dropout rates obviously will be affected by policies aimed at privatizing education, and intuition provides that poverty would be highly correlated with both. The following *ggplot()* and *summary()* **(Fig. 2.1)** visualizations support this.

Figure 2.1. Correlation of illiteracy and dropout rates.




```
Call:
lm(formula = illiteracy ~ school_dropout, data = project_data)

Residuals:
    Min      1Q  Median      3Q     Max
-2.5631 -0.8018 -0.0934  0.3335  3.4432

Coefficients:
               Estimate Std. Error t value Pr(>|t|)
(Intercept)      1.5082     0.5808   2.597   0.0172 *
school_dropout   0.9956     0.2819   3.531   0.0021 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

From the B_1 intercept, one may notice how a one-percent change in school dropout rates leads to a 0.99 percent change in illiteracy; almost one-to-one. Similar results were reported with poverty, so tests were conducted to determine the magnitude of omitted variable bias by disincluding *illiteracy* and *school_dropout*- poverty essentially captures the explanatory effect of all three, so to save *k* degrees of freedom for subsequent F-tests the intuition was to remove some of these three. Using the formula $OVB = \widehat{B2} * (cov(x1, x2) / var(x1))$ resulted in very tiny negative bias from omitting both; -0.022758 and -0.005014, respectively. The decision was made to only include *poverty* as a measure for educational attainment and accessibility.

Finally, the previously-discussed *buenos_control* dummy variable was tested for impact on the model. The following two *stargazer()* (**Fig 2.2)** layouts were produced.

Figure 2.2. Magnitude of control variable impact.

```
================================================
                         Dependent variable:
                     ---------------------------
                                gdp
------------------------------------------------
pop                           19.020***
                               (0.592)

Constant                    -1,517,283.000
                           (2,112,329.000)

------------------------------------------------

================================================
                         Dependent variable:
                     ---------------------------
                                gdp
------------------------------------------------
pop                           22.091***
                               (2.265)

buenos_control              47,972,028.000
                           (34,212,983.000)

Constant                    -52,488,110.000
                           (36,410,182.000)
```

The above tables demonstrate the increased magnitude of the *population* variable by controlling for Buenos Aires, also demonstrating a slight negative bias on the averages for population's explanatory influence when Buenos Aires is included in the model.

The result from all of this data work is the following explanatory variables:

logpop, logdoc, deficient_infra, poverty, no_healthcare, buenos_control.

Interpreted as: log(population), log(doctors per capita), percent lacking infrastructure, percent lacking healthcare, and the control dummy variable for Buenos Aires.

The following summary statistics visualization (**Fig 2.3**) was produced.

Figure 2.3. Summary statistics.

```
     loggdp          logdoc          logpop
 Min.   :15.15   Min.   :-5.870   Min.   :12.52
 1st Qu.:15.90   1st Qu.:-5.507   1st Qu.:13.15
 Median :16.21   Median :-5.348   Median :13.56
 Mean   :16.49   Mean   :-5.359   Mean   :13.72
 3rd Qu.:16.81   3rd Qu.:-5.234   3rd Qu.:14.02
 Max.   :19.49   Max.   :-4.588   Max.   :16.56
no_healthcare    deficient_infra    poverty
Min.   :29.23    Min.   : 3.84    Min.   : 3.399
1st Qu.:45.55    1st Qu.: 7.57    1st Qu.: 7.473
Median :49.37    Median :10.87    Median : 9.142
Mean   :50.77    Mean   :12.68    Mean   : 9.926
3rd Qu.:56.92    3rd Qu.:16.10    3rd Qu.:12.500
Max.   :65.81    Max.   :31.48    Max.   :17.036
```

IV. Methodology

The econometric procedure consisted of two multiple ordinary least squares (OLS) regressions, both logarithmically-transformed, and the second consisting of a control for Buenos Aires. The idea is to yield a GDP-explanatory elasticity model, which was accomplished through the following two regression models (**Fig 3.1, Fig 3.2**). The predicted effects are that population and doctors per capita will be positively correlated to output, while lack of access to healthcare, lack of infrastructure, and poverty will be negatively correlated. Furthermore, poverty, doctors per capita, and healthcare touch on a province's access to education and healthcare, while lack of infrastructure relates to government spending on public services. Note how these are all key talking points of Milei's policy proposals.

$$loggdp_i = \beta_0 + \beta_1 logpop_i + \beta_2 logdoc_i + \beta_3 deficient\_infra_i + \beta_4 no\_healthcare_i + \beta_5 poverty_i + e_i$$

$$loggdp_i = \beta_0 + \beta_1 logpop_i + \beta_2 logdoc_i + \beta_3 deficient\_infra_i + \beta_4 no\_healthcare_i + \beta_5 poverty_i + \beta_6 buenos\_control_i + e_i$$

Figure 3.1. First logarithmic OLS regression model.

	Dependent variable:
	loggdp
logpop	0.971***
	(0.087)
logdoc	0.010
	(0.377)
deficient_infra	0.008
	(0.013)
no_healthcare	-0.037**
	(0.013)
poverty	-0.048
	(0.032)
Constant	5.447**
	(2.407)

Figure 3.2. Second logarithmic OLS regression model, controlling for Buenos Aires.

	Dependent variable:
	loggdp
logpop	0.903***
	(0.132)
logdoc	0.135
	(0.424)
deficient_infra	0.009
	(0.014)
no_healthcare	-0.034**
	(0.014)
poverty	-0.045
	(0.032)
buenos_control	-0.365
	(0.528)
Constant	7.251*
	(3.576)
Observations	22
R2	0.924
Adjusted R2	0.893
Residual Std. Error	0.332 (df = 15)
F Statistic	30.219*** (df = 6; 15)
Note:	*p<0.1; **p<0.05; ***p<0.01

From the regressions, it is clear that the selected explanatory variables did a fantastic job at encapsulating the predicted percent change in GDP ($R^2 = 0.921$). Furthermore, by choosing to control for Buenos Aires, the magnitude of the R^2 increased to 0.924, meaning that 92.4% of the change in predicted GDP can be explained by the selected explanatory variables. Note that only *no_healthcare* and *logpop* were statistically significant at $\alpha < 0.05$, pointing to the importance of population and access to healthcare as a predictor for GDP.

Owing to the reported statistical insignificance of the remaining explanatory variables, an F-test (**Figure 3.3**) was conducted to determine whether the model remained significant with the inclusion of these variables.

Unrestricted model:

$$gdp_i = \beta_0 + \beta_1 logpop_i + \beta_2 logdoc_i + \beta_3 deficient\_infra_i + \beta_4 no\_healthcare_i + \beta_5 poverty_i + \beta_6 buenos\_control_i + e_i$$

Restricted model:

$$gdp_i = \beta_0 + \beta_1 logpop_i + \beta_4 no\_healthcare_i + \beta_6 buenos\_control_i + e_i$$

For

$$\mathbf{H_0: \beta_2 = \beta_3 = \beta_5 = 0}$$

$$\mathbf{H_1: \beta_2 \neq \beta_3 \neq \beta_5 \neq 0}$$

Figure 3.3. F-test for model significance.

```
  Res.Df    RSS Df Sum of Sq      F  Pr(>F)
1     18 2.4871
2     15 1.6540  3   0.83307 2.5183 0.09746 .
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

While a relatively-high p-value of 0.09746 resulted, it is still significant at $\alpha < 0.1$, which is sufficient enough to reject the null hypothesis in favor of the alternative hypothesis. In summation, we have sufficient evidence to believe that the provincial population, proportion of the population in poverty, proportion of the population without access to healthcare, proportion of the population without infrastructure, and doctors per capita are valid predictors of GDP in Argentina, on average.

V. Results.

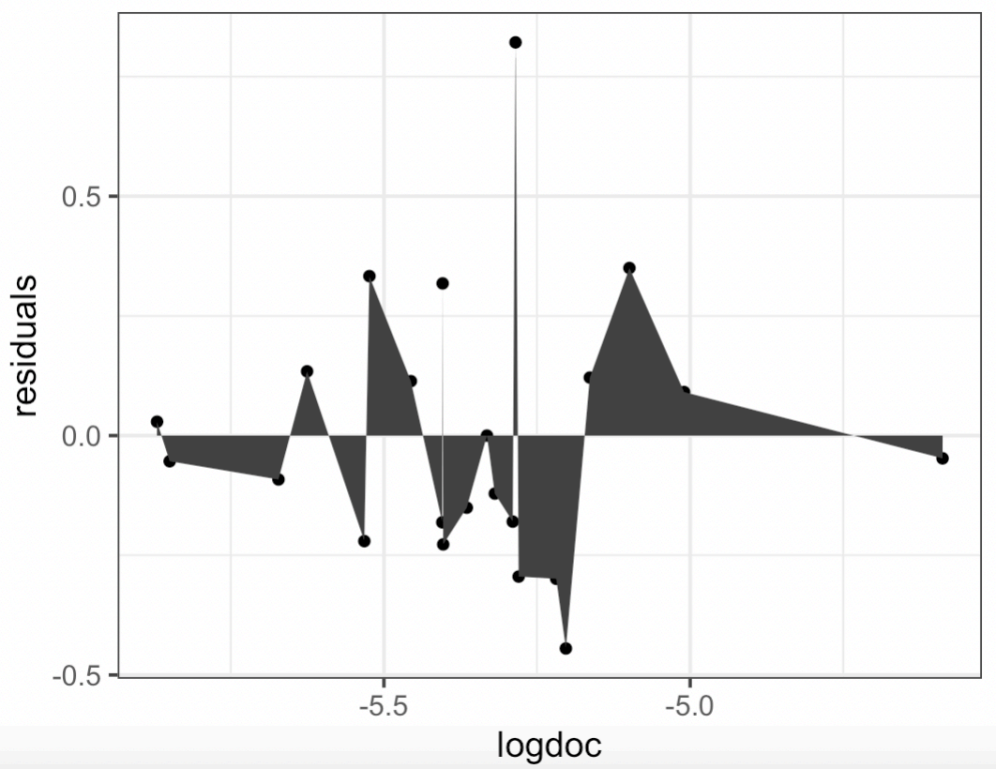
The final analysis of **IV. Methodology** revealed the validity of the chosen explanatory variables in predicting Argentine output, as well as relating their inclusion in the model to the purposes of analyzing Milei's fit for Argentine rule. One noteworthy quality of the second regression model is the changes in magnitudes brought about by controlling for Buenos Aires. While all explanatory variables had a minor change in magnitude, *logdoc*- the variable concerned with doctors per capita- became nearly thirteen times as strong in explanatory power when Buenos Aires was controlled for.

There are many reasons why the importance of the number of doctors per capita is so much greater outside Buenos Aires. It could be that in Buenos Aires, the average income is so much greater than outside provinces that there are many more pharmaceutical stores and non-practicing health clinics which decreases the importance of trained, practicing physicians. It could also be that Buenos Aires's position as an international city implies there are many tourists or foreigners with access to healthcare outside of Buenos Aires, once again decreasing the importance of doctors within Buenos Aires. If Buenos Aires was not controlled for, the model

would miss the very notable significance of the importance of doctors in affecting real output in provinces outside Buenos Aires- they likely don't have access to the same international acclaim and capital inflow that Buenos Aires does.

The last step in checking the performed OLS regressions as the best linear unbiased estimator (BLUE) was to perform heteroskedasticity tests. To thoroughly ensure homoskedasticity, a Goldfeld-Quandt Test and White Test were administered.
From a slew of *ggplot()* visualizations, the regression of *residuals* on *logdoc* (**Figure 4.1**) appeared the most heteroskedastistic.

Figure 4.1. Alleged heteroskedasticity in doctors per capita.



Surprisingly enough, neither the Goldfeld-Quandt nor White tests reported sufficient evidence of heteroskedasticity. A Goldfeld-Quandt test statistic of 9.138 led to a p-value of 0.203385, much higher than the threshold value for statistical significance. The White test

yielded a LaGrange Multiplier of 22, which produced a p-value of 0.7813- once again, much higher than the statistical significance rejection region. Therefore, there is not enough evidence to conclude that the data is subject to heteroskedasticity, and the regression coefficients are valid in interpretation. Something interesting to note is the ridiculous length of the White test (**Fig 4.2**) regression from having seven parameters.

Figure 4.2. A very tedious regression.

```
# white test
r_white <- resid(secondreg)
resid_reg <- lm(
  I(r_white^2) ~ logpop + logdoc + deficient_infra + no_healthcare +
    poverty + buenos_control
  + I(logpop^2) + I(logdoc^2) + I(deficient_infra^2) +
    I(no_healthcare^2) + I(poverty^2) + I(buenos_control^2) +
    logpop:logdoc + logpop:deficient_infra +
    logpop:no_healthcare + logpop:poverty + logpop:buenos_control
  + logdoc:deficient_infra
  + logdoc:no_healthcare + logdoc:poverty + logdoc:buenos_control +
    deficient_infra:no_healthcare + deficient_infra:poverty +
    deficient_infra:buenos_control +
    no_healthcare:poverty + no_healthcare:buenos_control +
    poverty:buenos_control, data = project_data
   )
```

VI. Conclusion.

Based on the results from a lengthy econometric procedure, we have procured a great deal of statistically significant evidence that the main predictors for Argentine output are provincial populations and open access to healthcare, with doctors per capita and poverty in tow. Understanding this, we can begin our analysis of Milei's policy. Milei has made clear his intent to privatize healthcare, having already removed regulations for healthcare companies per his Emergency Decree in December 2023. Admittedly, it is unknown by the extent of this paper if privatizing healthcare will lead to a lesser accessibility of healthcare. According to economic intuition, however, privatizing healthcare will likely increase the premiums required to afford it, which in turn will *increase* the proportion of the population without access to healthcare; this increase is reported by the regression model as having a negative effect on output, therefore having a negative effect on the well-being of the economy. The same can be said about the number of doctors per capita; the number of doctors per capita is likely positively correlated with the accessibility of education. If privatizing education leads to a further decrease in its accessibility, thereby decreasing the number of doctors per capita, this will have a negative effect on Argentina's GDP. Any inference made about Milei's policy goals and their relation to population seem far-fetched; perhaps social unrest brought about by his controversial implementations will cause a mass exodus, further decreasing GDP, but this is completely unknown. Something similar can be said with the poverty explanatory variable, although it seems feasible that a lack of government safety nets will cause more citizens to plunge into poverty. The likes of Mises and Konkin would argue, however, that with less safety nets, people will be more motivated to innovate and participate in the labor market, which will actually decrease the proportion of the population in poverty.

This research paper was very limited in its scope. For one, the available macroeconomic indicators leave much of the policy analysis to assumptive statistical inference, basing my normative review of his policy on assumptions about the world rather than empirically-tested statistics. This could be improved with another year of data to see how prices of college and healthcare premiums change; essentially any data that I made assumptions on in the previous paragraph could be hypothesis tested and lead to evidence-based conclusions at statistically significant p-values, which would greatly legitimize the interpretations. Another change to improve the quality of the statistical analysis would be to find a county-level dataset. Provincial data utilizes only 22 ($n = 22$) observations, which makes things like F-tests and Chi-squared tests more difficult as adding parameters becomes very limiting to degrees of freedom with such a tiny amount of observations. To find county-level data would legitimize the findings of any hypothesis tests.

In conclusion, by the use of econometric procedure and multiple logarithmic OLS regression, we can conclude that access to healthcare, population, poverty, and the number of doctors per capita are, on average, the most important macroeconomic determinants for provincial real GDP growth in Argentina. In addition, we have found that Milei's policy choice to target certain areas which seem highly predictive for GDP growth, like healthcare and education, may be ultimately harmful for Argentina's capacity for real output growth. Finally, this is all tied together in importance as theoretical economic modeling demonstrates that real output can be considered the most important determinant of macroeconomic well-being. If Milei's policy targets areas which very well may decrease the real output, it can be said that, based on the statistical analysis and limited scope of this paper, Milei is an unfit ruler for Argentina.