# GV101 Intro to Polsci

## Professor Simon Hix

## Regression Revision Document

### Cedric Tan

### May 2019

**Abstract**

This is a review of regression tables and how to interpret them from Simon Hix's lectures in the academic year of 2018-2019. The notes are mine fully and may not be authentic to the lecturer's as they have been modified.

The format of this material is taken from lectures where necessary and classes with Dr Anastasia Ershova.

# Contents

# 1 Overview of Regression

Regressions can be run for several purposes including:

1. To give a **descriptive summary** og how the outcome varies with the explanatory variables

2. To **predict the outcome** given a set of values for the explanatory variables

3. To **estimate the parameters of a model** describing a process that generates the outcome

4. To **study causal relationships** that could be explained by the model

## 1.1 Descriptive Summary

Through regressions such as Ordinary Least Squares (OLS), we achieve the best-fitting linear relationship, where best is defined as minimizing the sum of squares of the residuals i.e. difference between predicted and real values.

OLS estimates the best-fitting line in the population. However, the summary provided by OLS may miss important features of the data, such as outliers or non-linear relationships.

## 1.2 Prediction

OLS regression gives the best linear predictor in the sample. If the sample is drawn randomly from a larger population, OLS is a consistent estimator of the population's best linear predictor.

## 1.3 Estimation and Causality

Estimating the parameters of a model is the purpose that receives the most discussion in traditional textbooks. However, causality is the real motivation for regression. This is causal inference.

# 2 Basic Terms for Regression

Below are some basic terms for regression that one should memorise to use at all times:

- **Direction:** this is the way the effect is going, either positive or negative which corresponds to an increase or decrease

- **Neutral Language:** ensure to use this language if you are not confident in the units that are associated with the change and the direction of the change. For example: *there is a unit increase in Y due to X*

- **Controls:** these are effects that have influence on other variables which can ultimately affect the magnitude and direction of the dependent variables

- $R^2$**:** this is a measurement of how much variance is explained by the dependent variables in the model

- **Variance:** helps understand the differences in the level of dependent variables that occur depending on other variables
- **Confidence:** this explains how much faith you have in the fact that something occurred by chance or is truly explained by the model constructed
- **Omission:** this is when a model has missed out crucial variables which might make insights biased one way or another

# 3 Language use for Regressions

When explaining certain parts of a regression table, there needs to be a critical understanding of the language used to answer how the table works. Below are these critical phrases to keep in mind:

- When testing for a single variable: remember that the answer given in the regression output is while **holding all other variables constant.**
- When comparing models, recognise that **bias might be present due to omission** which is a key factor when seeing the difference in the amount of variables present in one model over another
- Underspecification which means it suffers from omitted variable bias
- Absorption means taking in the effect of another variable which is highly correlated. When adding one variable, the other may become insignificant as a result of this absorption

# 4    Full Examples

| Example 1: | | |
|---|---|---|
| Variables | Model 1 | Model 2 |
| Authoritarianism | | 0.46 |
| | | (0.23)** |
| Crisis | 0.87* | 0.88 |
| | (0.51) | (0.54) |
| Inflation | | -0.50** |
| | | (0.02) |
| GDP Per Capita | -0.001** | -0.001*** |
| | (0.00) | (0.00) |
| Constant | 9.12 | 9.10 |
| R Squared | 0.12 | 0.20 |

*p>0.90; **p>0.95; ***p>0.99

**Dependent Variable:** Level of Unemployment
**Independent Variable:** Authoritarian Govt. (1 = Authoritarian)
**Controls:**

- Crisis (1 = year of crisis starting)

- Inflation (%)

- GDP per Capita (USD)

1. What is the effect of GDP on the level of unemployment in Model 1?

2. How does the effect of Crisis changes from Model 1 to Model 2? Why?

3. What does the change in R squared indicate?

4. How do the changes in the models affect our understanding of the results?

**Question 1:**
An increase in GDP by one US Dollar would lead to a 0.001 unit decrease, i.e. have a negative effect, in unemployment holding all other control variables constant. This result is statistically significant to the 95th percentile.

**Question 2:**
Crisis has a positive effect on the level of unemployment with a magnitude of positive 0.87 units per unit of crisis increasing. It is statistically significant to the 90th percentile. However, in model 2, the crisis effect becomes not statistically significant at all despite it increasing in magnitude by 0.01. This may be due to the omitted variable bias where the effect of inflation may have absorbed the effect of crisis due to their high correlation with inflation being a more influential factor to unemployment.

**Question 3:**
R squared measures the amount of variance which is captured by the dependent variables in the model. The change in R squared shows that the second model, which captures 0.08 units more, accounts for more of the variance associated with unemployment.

**Question 4:**
hello

**Example 2:** ☐