# **Research Proposal - Penn World Tables Data**

**What is your hypothesis?**

My hypothesis is that an increase the number of expected years of education leads to an increase in a country’s per-capita output-side real GDP in 2017 US Dollars. We will examine this hypothesis using panel data covering all 27 EU member states, as well as the United Kingdom, from 1970 to 2019.

H0: An increase in the number of expected years of education will not increase a country’s per-capita output-side real GDP in US Dollars.

H1: An increase in the number of expected years of education will increase a country’s per-capita output-side real GDP in US Dollars.

**What is your main baseline specification?**

My baseline specification will be a simple OLS regression to examine the relationship between number of years of education (independent variable) and per-capita output-side real GDP (dependent variable). The baseline model will be across all EU member states, as well as the United Kingdom, from 1970-2019.

After examining the results of this simple regression, I may need to transform the dependent variable to address issues such as non-linearity, or lagged effects, between years of education and GDP output. If statistical testing implies need for a lagged effects component, I will conduct a short literature review to better understand what lag length may be appropriate for assessing the effect of change in number of years in education on output.

I will also improve the simple OLS regression approach taken in my baseline specification by testing for the appropriateness of alternative model types, such as a fixed effects model.

We will also add control variables, to account for other factors that may have influenced GDP output over the period shown. This is discussed further in the assumptions section.

**What data do you plan on using?**

I plan on using panel data from Penn’s World Tables (10.01), which shows country-level GDP metrics (dependent variable) from 1950 to 2019, and the World Bank’s World Development Indicators and Education Statistics datasets (independent variables), which range from 1970 to 2020. We will join these datasets, using data between 1970 and 2019.

My analysis will focus on the 27 EU member states, as well as the United Kingdom.

**What assumptions do you need to make in order for your approach to be a good test of your hypothesis?**

I have assumed that the relationship between years in education and GDP output per capita are relatively constant across different countries and years (in our fixed effects model). There may in fact be confounding variables that need to be accounted for as model development progresses. These might include: Unemployment rate, Net Exports, and Net Foreign Direct Investment figures.

I also assume that an increase in GDP per capita does not lead to an increase in number of years in education, which may not hold.

We also assume a set of classical assumptions regarding our simple regression approach:

1. There is no relationship between our error term and our explanatory variables (zero conditional mean).
2. Our model parameters/coefficients must be of the form (linear functional form).
3. Our error term has constant variance, regardless of the value(s) of the independent variable(s) (homoscedasticity).
4. Our error terms are uncorrelated with one another (zero error term correlation).

We do not have to assume that our data is representative of our population, as our panel data represents the population that we are interested in researching.

**What further strategies might allow you to relax some of these assumptions?**

It may be useful to include a lagged GDP variable in the set of predictor variables, rather than attempting to model GDP against the year in which changes to years in education occur. This assumption might impact the robustness of our final model, so we would need to find statistically significant evidence to retain the position that there is no lag between changes in educational policy and GDP output.

In our initial models, we have assumed that the model specifications provided accurately represent the relationship between different economic variables and GDP output, such that we can interpret the education-output relationship shown as causal. Instrumental Variables can control for confounding factors between number of years in education and GDP output and improve the robustness of this assumption.

**What policy implications may come from your results?**

This research aims to examine the relationship between number of years in education and per-capita output-side real GDP, in order to examine the hypothesis that an increase in the number of expected years of education will increase a country’s per-capita output-side real GDP in US Dollars.

Should this research find robust and statistically significant evidence of a relationship between years in education and GDP output, this will help to inform policymakers when examining or appraising education policy options. This might include the Lifelong Learning (Higher Education Fee Limits) Bill, Schools Bill, and other potential future education bills.