ECON3030

Numerical Methods in Economics

Dr. Juan Ignacio Vizcaino

Autumn 2024

Individual Coursework Instructions

Submit your coursework in a single Jupyter Notebook file (.ipynb) using the coursework submission link in the module's Moodle page. Follow the submission instructions in the submission link. Your code must run in my computer. Marks will be discounted each time code does not run. I expect individual and original answers, otherwise you should expect mark deductions, or the penalties stipulated in the University's academic misconduct policy. The use of AI tools is not allowed in assessments and is considered false authorship in the University's Academic Misconduct Policy (see the following link).

Introduction

The goal of this exercise is to update the Development Accounting exercise in Caselli (2003) using the latest version available of the Penn World Table (PWT 10.1).

Motivation

It is widely known that differences in standards of living across countries are enormous. Development accounting uses cross-country data on output and production inputs, at one point in time, to assess the relative contribution of differences in factor quantities and differences in the efficiency with which those factors are used to the vast differences in per-worker incomes. It is the same idea as growth accounting, but in the case focusing on income differences across space rather than time.

Conceptually, development accounting can be thought of quantifying the relationship

$$Income = F(Factors, Efficiency)$$

If one found that factors can account for most of the differences, then development economics should focus on explaining low rates of factor accumulation. If efficiency differences play a large role instead, then more research is needed to understand why some countries extract more output than others from their factors of production.

Operationally, they key steps in development accounting are:

1. Choosing a functional form for F, which is assumed to be **common across countries**.
2. Accurately measuring Income and Factors.
3. Back out efficiency as a residual. As in the case of the Solow residual, this residual is a measure of our ignorance on the causes of poverty and under-development.

As in Caselli, we follow Hall and Jones and assume that a country's GDP (denote by Y) is produced using the following production function

$$Y = AK^{\alpha}(E \cdot avh \cdot h)^{1-\alpha}$$

where K is the aggregate capital stock and $L = E \cdot avh \cdot h$ is the "quality adjusted" workforce, namely the number of workers employed E multiplied by their average hours worked in a given year avh and by the average human capital per worker h. α is a constant that measures the share of capital income in total GDP, while $(1 - \alpha)$

measures the share of labour compensation in GDP. A represents the efficiency with which factors are used, or total factor productivity (TFP).

Since F exhibits constant returns to scale, it follows that in per-worker terms (i.e. $y = \left(\frac{Y}{E}\right)$) the production function can be written as

$$y = Ak^{\alpha}(avh \cdot h)^{1-\alpha},$$

where k is capital per worker $k = \left(\frac{K}{E}\right)$, y is output per worker $y = \left(\frac{Y}{E}\right)$, h and avh are average hours worked and average human capital per worker, respectively. ***We want to know how much of the variation in y can be explained by the variation in the observables, avh, h and k, and how much is residual variation*** *(i.e. that can be attributed to differences in A)*.

Data

In his paper Caselli uses data from PWT 6.1 (Heston, Summers, and Aten (2002)). More than twenty years have passed and the PWT has evolved. The current version, PWT 10.1 [1], introduces significant changes that improve the quality of the data and the variables available, facilitating the development accounting exercise.

In what follows, we will measure:

- **Aggregate Output** (Y): using PWT's *rgdpo* (*Expenditure-side real GDP at chained PPPs (in mil. 2017US$)*)
- **Employment** (E): using *emp* (*Number of persons engaged (in millions)*)
- **Hours Worked** (avh): using *avh* (*Average annual hours worked by persons engaged*)
- **Human Capital** (h): using variable *hc* (*Human capital index*).
- **Share of labour compensation in GDP** ($1-\alpha$): using the variable *labsh (Share of labour compensation in GDP at current national prices)*.
- **Total Factor Productivity** (A): depending on the exercise, we will either use the variable *rtfpna* (*TFP at constant national prices (2017=1)*) or compute this variable residually.
- **Capital Stock** (K): using PWT's *cn* (*Capital stock at current PPPs (in mil. 2017US$)*)

Questions

1. Preliminaries (15 points)
    a. Import the PWT 10.1 to Python using the *pandas* library. Use the following links to my GitHub repo where I host both a .csv file (link .csv) and an .xlsx file (link .xlsx). Call your dataframe pwt101.

    b. Filter your dataframe to select only the variables of interest, as described above.

    c. Choose a year to perform your development accounting analysis. The year chosen must be the latest year available that guarantees having the maximum number of observations possible to perform your development accounting results. *Hint 1: when working with dataframes only countries that have observations for all variables of interest will be used in your calculations.*

    d. Provide a table of descriptive statistics characterizing output per worker ***in your subsample***. Include mean, standard deviation, max, min, and percentiles 5,10,25,50,75,95. Tabulate your results accordingly. Add a column showing the ratio of output per worker of your statistics with respect to the United States. The first two columns of your table should include the country and countrycode corresponding to each of these statistics (but the variance). *Hint: here you need to find just one country per statistic.*

    e. Provide a table of descriptive statistics with the same structure and content as in question 1.d, but now including all countries with data on output per worker for your chosen year. What conclusion can you draw when you compare the results obtained in questions *1.d.* and *1.e.*?

[1] See the PWT 10.1 website in the following link for a description of the Penn World Table and documentation. In case it helps, you can also access the data in Excel and Stata format, but your work must be done in Python using the instructions provided in this file.

2. In this exercise we will inspect our variables of interest graphically to analyse their behaviour across the development spectrum. (10 points)
   a. To do so, you will produce a series of scatterplots showing $\ln(y) = \ln\left(\frac{Y}{E}\right)$ in the x-axis against $\ln(k)$, $\ln(h)$, avh, and $(1-\alpha)$ in the y-axis, respectively. Label the axes accordingly. Add a title to each of your charts. Add a regression line together with its corresponding equation to each chart. Add vertical dashed lines to your chart denoting the 25th,50th, and 75th percentiles of $\ln(y)$.

   b. What conclusion can you draw from this graphical analysis?

A Basic Mesure of Success

Set $(1-\alpha)$ to the average value of *labsh* in your sample. With data on k, h, y, avh together with the functional form defined above, we can compute output per worker as

$$y = \tilde{A} \cdot k^{\alpha} \cdot (avh \cdot h)^{1-\alpha},$$

where $\tilde{A}$ is residual TFP. To be precise, we can rewrite the production function in per-worker terms as $y = \tilde{A} \cdot y_{kh}$, with $y_{kh} = k^{\alpha} \cdot (avh \cdot h)^{1-\alpha}$. Notice that both y and y_{kh} are measurable using PWT data, and that $\tilde{A}$ can be compute residually as $\tilde{A} = \frac{y}{y_{kh}}$.

Having computed $\tilde{A}$ and y_{kh}, we can proceed to perform our development accounting exercise. The goal is to answer the following question: Suppose that all countries had the same level of efficiency $\tilde{A}$; what would the world income distribution look like in that case, compared to the actual one? To perform this assessment, we will look at a few alternative measures:

- *Measure 1:* The first one is in the tradition of variance decomposition and shows what share of the variance in income per-worker can be explained by factors of production.

$$success_1 = \frac{var[\ln(y_{kh})]}{var[\ln(y)]}$$

- *Measure 2:* We are also interested in measuring the share of the variance in income per-worker that can be explained by residual TFP, which can be computed as

$$success_2 = \frac{var[\ln(\tilde{A}\,)]}{var[\ln(y)]}$$

- *Measure 3:* Additionally, we are interested in knowing what share of the variance in income per worker can be explained by the measure of TFP provided by the PWT, defined by A (using variable *rtfpna*)

$$success_3 = \frac{var[\ln(A\,)]}{var[\ln(y)]}$$

- *Measure 4:* A final measure, less sensitive to outliers, compares inter-percentile differentials. In the case of the 90th and 10th percentile this statistic is given by

$$success_4^{90,10} = \frac{\frac{y_{kh}{}^{90}}{y_{kh}{}^{10}}}{\frac{y^{\ \ 90}}{y^{\ \ 10}}}$$

  where $y^{\ \ 90}$ and $y^{\ \ 10}$ are the 90th and 10th percentile of our measure of GDP per worker.

3. Compute these measures of success in your PWT sample. (15 points)
   a. As in Caselli, report these results in a table including the resulting measure of success, the number of observations used to compute it, and the name of the measure computed in each case.

b. Can you think of any other measure of success? Compute your new measure and add it to the table provided above.
c. What do you conclude from this analysis?

4. Keeping α fixed in the value computed above, we can decompose $var[\ln(y_{kh})]$ as $var(ln(y_{kh})) = \alpha^2 var(ln(k)) + (1-\alpha)^2 var(ln(avh \cdot h)) + 2\alpha(1-\alpha)cov(ln(k), ln(avh \cdot h))$. (15 points)
    a. Compute what share of the $var(ln(y_{kh}))$ that is explained by each of the terms in the right-hand side of the equation above defining $var(ln(y_{kh}))$. Report your results in a table. *Hint: the contribution of each term is its value divided by* $var(ln(y_{kh}))$.
    b. What do you conclude from this analysis?

5. Based on your calculations: are differences in standards of living across countries mostly driven by factor accumulation or by the efficiency in which factors are used? Do your results depend on the measure of success used? Which factor of production is more important in explaining cross-country differences in income per-worker? How do your results compare to those in Caselli (2003)? (10 points; Word limit: 350 words).

6. Can you think of any missing factors explaining differences in income per worker that we have not accounted for in this exercise? Provide a quantitative assessment of these factors either adding additional variables to your analysis or by modelling these factors accordingly. *Hint6: For example, in section 4.4 Caselli argues that cross-country differences in health and nutrition status also play a role in explaining human capital differences. He models these effects in a simple, reduced-form way, and provides data to quantitatively assess the role of health gaps in explaining income differences across countries.* (15 points)

7. What conclusion can you draw from the exercise conducted? Are there any policy recommendations that emerge from the exercise? How do your results compare to the existing literature? Are there any avenues for research that come out of the exercise? (20 points; Word limit: 500 words)

References

- Caselli, F. (2005) Accounting for cross-country income differences. In P. Aghion & S. N. Durlauf (eds.), Handbook of Economic Growth, Vol. 1A (pp. 679–741). Amsterdam: Elsevier
- Hall, R.E., and C.I. Jones, (1999), “Why do some countries produce so much more output per worker than others?”, The Quarterly Journal of Economics 114(1):83-116.
- Heston, Alan; R. Summers, and B. Aten (2002), Penn World Tables Version 6.1. Downloadable dataset. Center for International Comparisons at the University of Pennsylvania.
- Feenstra, Robert C., Robert Inklaar and Marcel P. Timmer (2015), "The Next Generation of the Penn World Table" American Economic Review, 105(10), 3150-3182, available for download at www.ggdc.net/pwt