Research Assistantship Coding Task

Samuel Suárez Valle, Universidad de los Andes

August 2025

Contents

1 Simple data Analysis

1.1 Setting up Github

This coding task is stored in the following **Github repository**:
https://github.com/SamuelSuaV/CodingTask-SamuelSuarez

1.2 Database description

According to the User's Guide to PWT 10.01, the Penn World Table (PWT) 10.01 is a panel data set that provides annual data for 183 countries spanning the period from 1950 to 2019. It includes data on **GDP** and its components across various measurement approaches; **external sector** indicators such as exports and imports; **labor** market variables including total labor force, human capital, and labor income share; and **capital**-related data such as the capital stock, depreciation rates, and investment. Furthermore, the data set facilitates international comparisons through purchasing power parity (PPP) and standardized expenditure data, as compiled by the International Comparison Program (ICP).

After an overview of some papers where the PWT data set has been cited, we see that some types of analysis that this dataset allows include **time series** regression (Ahmed et al., 2020), **panel** regression (Sun et al., 2021; Caglar and Askin, 2023), as well as **cross-sectional** regressions if the data is limited to one period.

1.3 Cleaning up the data

The PWT 10.01 database was cleaned by keeping the relevant variables between the chosen time period (1990-2019). These variables include the **year**, **output-side GDP** at current PPPs (in million 2017 USD), the **population** in millions, **GDP per capita**, the **total factor productivity** in current PPPs with the USA productivity equal to 1, and the **labor share**. In the list below, you can see why some of these variables were chosen to be kept in the database, while other measures of the same variable were left out. The variables left out of this list were not available in different forms of measurement.

- **Output-side GDP**: This variable was kept in units of millions of 2017 USD, as GDP values tend to extend to various orders of magnitude, making it simpler to analyze when kept in millions.
- **Population**: Like GDP, this variable tends to reach high orders of magnitude, for which analysis in millions becomes easier.
- **GDP per capita**: Unlike the previous two variables, GDP per capita ranges between smaller values, which makes analysis in units more convenient.
- **Total factor productivity**: This variable was kept in constant PPPs to make it comparable with output-side GDP, which is only available in

constant PPPs. Additionally, while standardizing the USA productivity to 1 limits our ability to analyze growth trends, this allows us to compare other countries to the frontier more conveniently.

1.4 Creating new variables

A series of new variables where added to the database for analysis. In particular, I included the **first difference in logarithms** of per capita GDP and total factor productivity; the **annual growth rate** of GDP per capita, total factor productivity and the labor share; and the **cumulative growth rate** of GDP per capita, total factor productivity and the labor share.

Logarithmic first-difference variables (computed as $[\ln(x_t) - \ln(x_{t-1})] \times 100$) are used as an approximation of annual percentage changes in each variable. Unlike simple first differences, logarithmic first differences represent proportional changes such as growth rates and are scale-invariant, which makes them suitable for growth analysis, where simple first differences would become harder to interpret as more variables with varying units are included. Logarithmic first differences are also easier to decompose, as adding logarithmic first differences across k periods is equivalent to calculating the log differences between the range of k periods.

On the other hand, annual growth rates (computed as $\frac{x_t - x_{t-1}}{x_{t-1}} \times 100$) measure actual growth rates, rather than approximated ones. Percentage change measures offer the same advantages of scale invariance as logarithmic first differences, however, they also allow for simpler interpretation as compared to analysis of logarithms. Additionally, while first logarithmic first differences tend to approximate percentage changes for lower growth rates, these two tend to diverge for larger growth rates as seen in **Figure 1**.

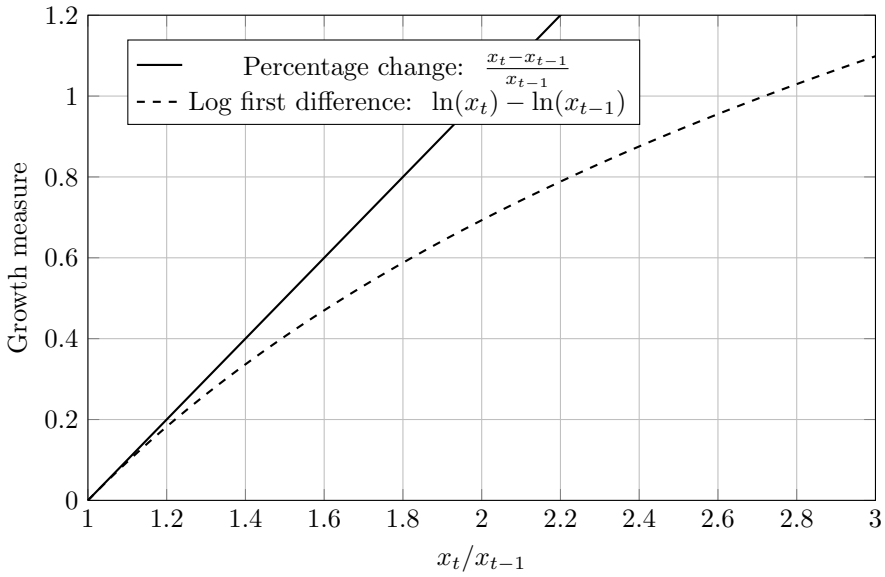

Figure 1: Comparison between percentage change and log first difference
Note: Figure produced with assistance from OpenAI's ChatGPT.

Finally, cumulative growth rates (computed as $\left(\frac{x_t}{x_{1990}} - 1\right) \times 100$) are used to measure the growth of a variable from a specific time period, rather than within two consecutive time periods, hence their differences in values. These growth rates can be used to measure long term scale invariant changes to variables, rather than instantaneous ones.

1.5 Saving the dataset

The finished dataset was saved in the following Github file path:
$CodingTask - SamuelSuarez/Data/Processed/build_dofile.dta$

1.6 Time series plots

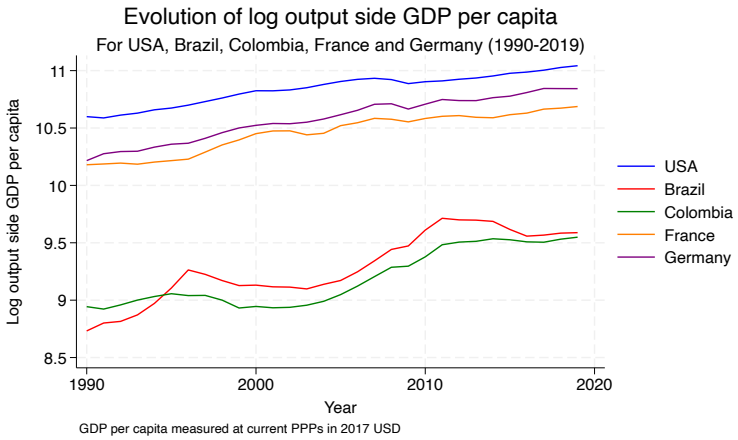

Figure 2: Evolution of log output-side GDP per capita (1990–2019).

Figure 2 shows the yearly evolution of log per capita GDP for the five analyzed countries since 1990 until 2019. The graph shows that, although growth trends for the South American countries are relatively similar to those countries in the north, Colombia and Brazil present an overall lower GDP per capita throughout this period. The South American countries also present similarities in their growth patterns with stagnation periods beginning in the late 90's followed by accelerated growth throughout the middle 2000's throughout the earlier years of the 2010's. These accelerated growth rates have now receded once again until 2019. On the other hand, the European countries, as well as the USA, have seen mostly constant growth throughout this period.

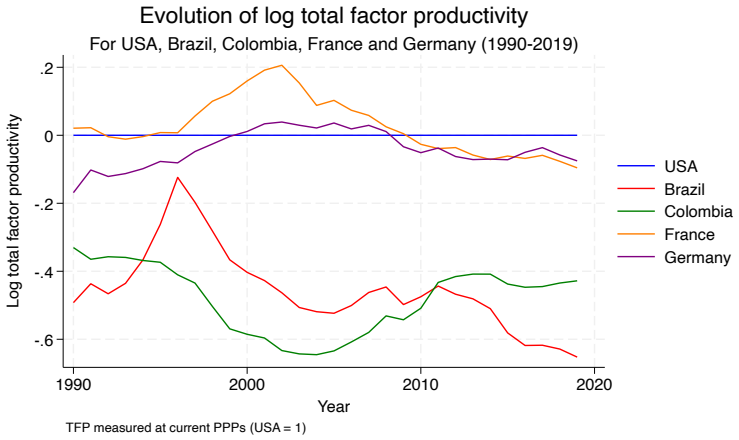

Figure 3: Evolution of log total factor productivity (1990–2019).

As mentioned earlier **Figure 3**, which shows log factor productivity trends, is limited in our capacity to analyze growth trends in factor productivity as USA productivity is standardized to 1. Relative to the US, however, one can notice that total factor productivity has stayed below the frontier for most countries throughout the analyzed period with the exception of France and Germany, which saw an increase in relative productivity starting in the late 90's which allowed for this countries to maintain a higher productivity than the US through most of the next decade. France highlights itself as the country which achieved the highest productivity throughout this period. In comparison, the South American countries, which already had lower productivity rates, saw a reduction in their relative factor productivity throughout the 2000's. After this period, Colombia recovered some of its losses in relative productivity while Brazil kept seeing a shrinking pattern. The case for Brazil stands as particularly interesting as this country saw standout growth in productivity throughout the late 90's, before shrinking rapidly.

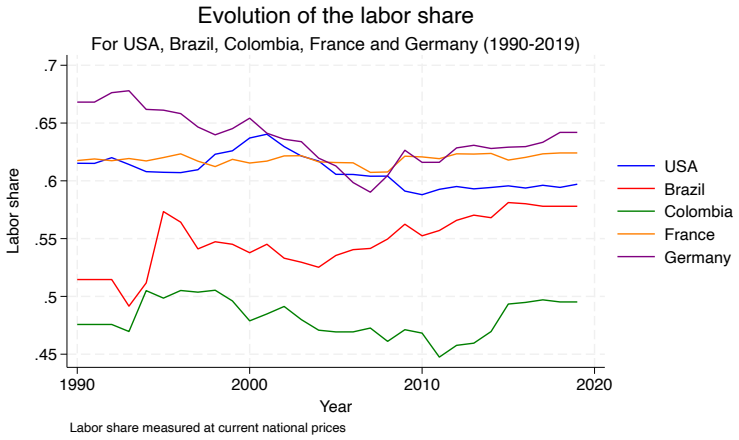

Figure 4: Evolution of the labor share (1990–2019).

Regarding **Figure 4**, which shows the evolution of the labor share in the analyzed period for the 5 selected countries, one can initially notice that, unlike the usual assumption of constant shares, labor shares tend to fluctuate rapidly even in singular time periods for all analyzed countries. The aforementioned is consistent with analysis from Kahn and Lim (1997) and Blanchard (1997), who also have measured factor shares in different periods to vary. Notice that the South American countries present a stagnated labor share through the late 90's, however Brazil has outgrew the Colombian share in the earlier 90's. Since then, the Colombian labor share has seen a reduction until the early 2010's, where the share grew back to previous similar levels, while the Brazilian share has seen growth through this period. Regarding northen countries, the US and France have seen more stable shares, with particular of growth and contraction for the United States around the year 2000. Germany displays higher variance in its share which shrunk until the middle 2000's before seeing low growth rates in the earlier 2010's.

1.7 Cumulative growth plots

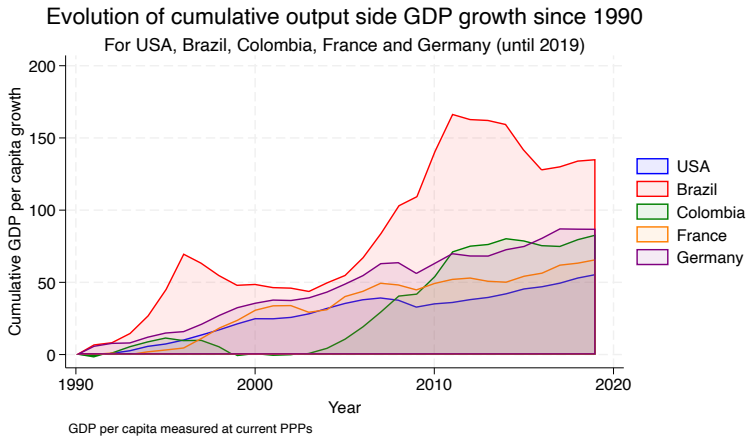

Figure 5: Cumulative growth of log output-side GDP per capita since 1990).

Figure 5 displays cumulative growth in the five selected countries from 1990 for each period until 2019. Brazil has seen by far the largest cumulative growth with peaks starting in the middle 1990's and the middle 2000's. Colombia, on the other hand, saw little (and even negative) cumulative growth rates until the middle 2000's where growth accelerated, with the country even showing the largest relative growth rate in the early 2010's. The US, France and Germany have seen stable growth rates since 1990 until 2019.

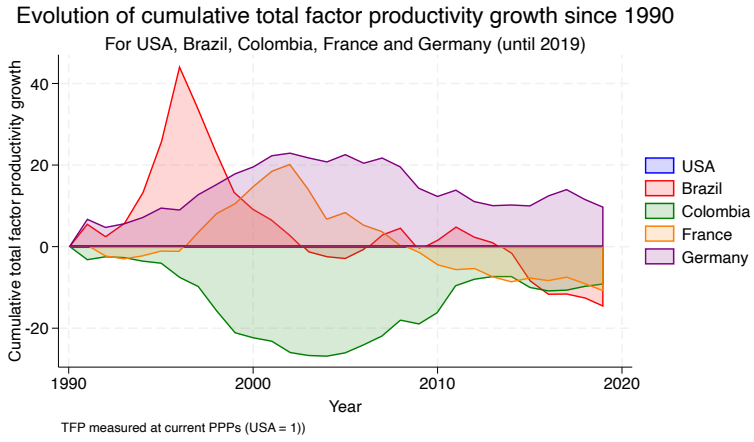

Figure 6: Cumulative growth of log total factor productivity since 1990.

Regarding factor productivity, **Figure 6** displays cummulative growth since 1990's for each period for Brazil, Colombia, France and Germany relative to the US. Since the 1990's, Brazil saw the highest variance showing a comparatively large peak in the late 1990's which lasted until the middle 2000's, where total factor productivity stagnated around the levels seen in 1990. Productivity for Brazil then shrank below the 1990 figures through the middle 2010's. Colombia, on the other hand, saw a continuous shrinking compared to the US in its factor productivity relative to 1990 until the middle 2000's, where productivity rose, although it never reached its original levels. France, on the other hand, saw stagnated growth rates through until the late 1990's where growth relative to 1990 became noticeable, compared to the US. By the late 2000's, this peak reached its end as productivity compared to the US turned lower than what it was in 1990.

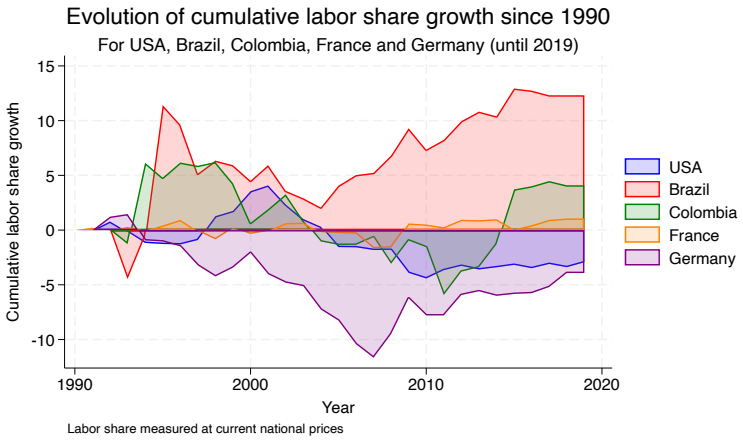

Figure 7: Evolution of the cumulative growth in labor share until 2019.

Figure 7 displays the cumulative labor share growth of all five countries relative to 1990. Notice that all countries exhibit variance around the comparison period, specially through the middle 1990's. After this period, Brazil displays the highest growth since 1990, beginning with a sharp rise after the middle 1990's before being reduced through the late 1990's until the late 2000's. Colombia, France and the US, on the other hand, exhibit smaller positive and negative growth rates around the 1990 values, with the US seeing a relative decrease in its labor share past the early 2000's. Opposite to Brazil, Germany displays a continuous shrinking in its cumulative labor share growth since 1990 with a convergence towards its original value past the late 2000's.

1.8 GDP vs TFP cumulative growth

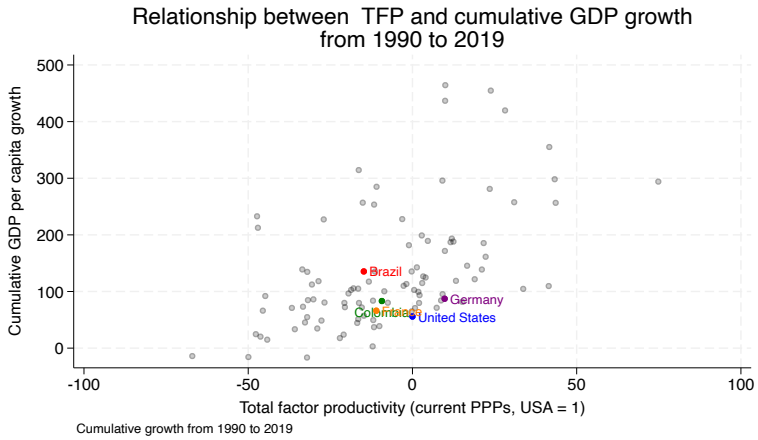

Figure 8: Relationship between cumulative TFP and GDP per capita growth from 1990 to 2019.

Figure 8 displays a scatter plot with cumulative GDP per capita growth until 2019 on the y axis and cumulative growth in TFP since 1990 on the x axis. Extreme values where omitted. The graph suggests the existence of a positive relationship between both variables, with the selected countries being in the middle of the plot. Brazil stands out as having a higher relative GDP growth than a factor productivity growth, while the US portrays the opposite pattern. The plot suggests that growth in factor productivity is closely related to growth in GDP per capita, as expected from classic literature.

1.9 Labor share vs GDP cumulative growth

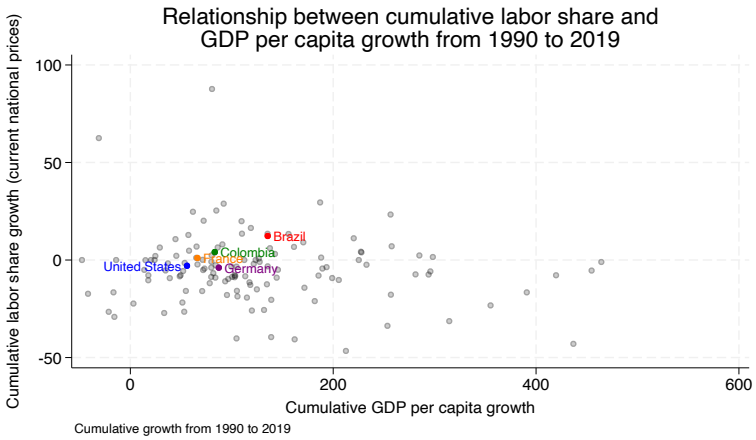

Figure 9: Relationship between cumulative labor share and GDP per capita growth from 1990 to 2019

Figure 9, on another hand, displays cumulative growth rates from 1990 to 2019 for the labor share, in the y axis and for per capita GDP in the x axis. Once again, extreme values where omitted. The United States sees nearly the lowest of both growth types in this periods while Brazil sees the most growth in both axes. Although it is unclear wether there is a pattern in this plot, in broad strokes, one can notice that GDP growth is positively associated with labor share growth.

References

Ahmed, Z., Asghar, M. M., Malik, M. N., and Nawaz, K. (2020). Moving towards a sustainable environment: The dynamic linkage between natural resources, human capital, urbanization, economic growth, and ecological footprint in china. *Resources policy*, 67:101677–.

Blanchard, O. J. (1997). The medium run. *Brookings Papers on Economic Activity*, 28(2):89–158.

Caglar, A. E. and Askin, B. E. (2023). A path towards green revolution: How do competitive industrial performance and renewable energy consumption influence environmental quality indicators? *Renewable Energy*, 205:273–280.

Kahn, J. A. and Lim, J.-S. (1997). Skilled labor – augmenting technical progress in u.s. manufacturing. Research Paper 9738, Federal Reserve Bank of New York.

Sun, H., Edziah, B. K., Kporsu, A. K., Sarkodie, S. A., and Taghizadeh-Hesary, F. (2021). Energy efficiency: The role of technological innovation and knowledge spillover. *Technological Forecasting and Social Change*, 167:120659.