Tax-Equity-in-Low-and-Middle-Income-Countries-Replication*

Amie Liu, Wanling Ma

February 16, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

You can and should cross-reference sections and sub-sections. We use R Core Team (2023) and Wickham et al. (2019a).

The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Source

The original paper, "Tax Equity in Low- and Middle-Income Countries," utilized for replication, was published in the American Economic Association journal in 2024 (Bachas, Jensen, and Gadenne 2024). It explores the role of taxation in mitigating income inequality within countries with lower and middle incomes (Bachas, Jensen, and Gadenne 2024). This paper seeks to build upon the findings of the original paper, using the same datasets that were last retrieved on February 10th, 2024.

The two main research questions that were explored in this paper: TODO

*Code and data are available at: https://github.com/WanlingMa/Tax-Equity-in-Low-and-Middle-Income-Countries-Replication

2.2 Methodology

This paper used the R programming language (R Core Team 2023) and the 'tidyverse' (Wickham et al. 2019b) packages for replication, reproduction, data cleaning, and analysis. Within the 'tidyverse' package (Wickham et al. 2019b), we used 'readr' (Wickham, Hester, and Bryan 2024), 'dplyr' (Wickham et al. 2023), 'tidyr' (Wickham, Vaughan, and Girlich 2024), and 'ggplot2' (Wickham 2016) for reading CSV files, manipulating data, tidying data, and creating graphs.

Data collection methods exhibit significant diversity among open sources. We conducted a data-cleaning procedure for reproduction, focusing on variable selection, simplifying variable names, and removing missing values. The variables selected from the original paper were chosen based on their relevance to our research question, relating to (TODO).

This paper analyzes the datasets listed below, which we have organized into three categories: Economic Development and Growth, Labor Market and Employment, and Taxation and Income Distribution.

2.2.1 Economic Development and Growth

`data/gdp_populaPon_WDI.dta`

The dataset was obtained from the World Bank's World Development Indicators Bank (2018). It encompasses a comprehensive sample from 266 countries and spans from the year 1960 to 2020 (Bank 2018). Three key variables available within the database are GDP in 2010 USD (gdp_2010usd), GDP in current USD (gdp_currentusd), and Population (pop). GDP per capita (USD) is calculated as the ratio of gdp_2010usd to population, providing a measure of economic output per person adjusted for inflation (Bank 2018). Meanwhile, GDP per capita (current) is calculated by dividing gdp_currentusd by population, offering a snapshot of economic output per person in current currency values (Bank 2018). This enables researchers to analyze economic growth, trends, and developments over an extensive time across diverse geographical regions.

`data/ross_mahdavi.dta`

The dataset was obtained from the Harvard Dataverse Ross and Mahdavi (2015). It encompasses information about oil and natural gas production of all countries from 1932 to 2014 (Ross and Mahdavi 2015). It contains six column variables, including the country ID (country), year, approximate value of oil and gas product in US dollars (natres_richesse), GDP worldwide (gdp_wid), GDP in constant 2014 US dollars (gdp_2014usd), and oil and gas product as percentage of GDP (oil_pct) (Ross and Mahdavi 2015). Covering the period from 1932 to 2014, this dataset provides a longitudinal perspective on oil and natural gas production trends with a global scope, providing valuable insights into the economic aspects of oil and natural gas production.

`data/globalETR_bfjz.dta`

The dataset was sourced from the paper titled "Globalization and Factor Income Taxation" (Bachas et al. 2022a) and is accessible through the website associated with the paper (Bachas et al. 2022b). It provides data on economic indicators and taxation in each country from 1965 to 2018, including the net domestic product in constant 2019 US dollars, effective tax rate, tax revenue, total income tax, social contributions, property and wealth tax, and indirect tax (Bachas et al. 2022b). This allows the authors to comparatively analyze the effective tax rate and taxation policies that influence income distribution globally.

`data/country_frame.dta`

The dataset was obtained from the World Bank's World Development Indicators Bank (2023). It includes several columns containing key indicators related to economic, social, and environmental development. These columns cover variables such as country codes represented by both two-letter (iso2) and three-letter (iso3), country names (country_name), region names (regionname), and income level classifications (incomelevelname) (Bank 2023). The income level classification consists of five distinct variables: high income, low income, upper middle income, lower middle income, and unclassified (Bank 2023). These variables provide insights into the economic, social, and environmental characteristics of countries worldwide.

2.2.2 Labor Market and Employment

`data/API_SL.EMP.SELF.ZS_DS2_en_csv_v2_5560396.csv`

The dataset was sourced from the International Labour Organization ILOSTAT (2021) and contains information from 1960 to 2022. It includes essential variables such as country code, country name, indicator name, indicator code, and the percentage of total employment each year (ILOSTAT 2021). By focusing on self-employment as a specific indicator type, the dataset enables the analysis of the proportion of self-employed people among the total workforce across different countries and over time.

2.2.3 Taxation and Income Distribution

`data/regressions_output_central.dta` `data/Country_informaPon.xlsx`

These regression datasets were obtained from the authors' article "Informality, Consumption Taxes, and Redistribution" Bachas, Gadenne, and Jensen (2023). The column variables include a unique identifier for each country (country_code), the year in which the data was collected (year), the number of iterations performed during the regression analysis (iteration), regression coefficient (b), standard error of the regression coefficient (se), and the coefficient of determination (r2_adj) (Bachas, Gadenne, and Jensen 2023). These were conducted for the regression analysis of informal shares in household consumption data.

`data/PIT_parameters_AJ.dta`

The dataset originates from Jensen's journal article "Employment Structure and the Rise of the Modern Tax System" Jensen (2022). It consists of data spanning from the year 1870 to 2014 and covers 100 countries. It includes several column variables, such as country code value (country_code), country name, year, mean of GDP per capita (lg_gdppc), mean of the marginal income tax rate (mtr), and mean of the size of personal income tax (size_pit) (Jensen 2022). These variables delve into the relationship between country-level exemption threshold, tax system evolution, and economic development across different countries and times.

`data/PIT_Top_Rates_2022.csv`

The dataset was collected and coded by the authors in 2023, providing information on the country-level top marginal tax rate of personal income tax for the year 2022 across various countries (Bachas, Jensen, and Gadenne 2024). This dataset provides insights into the taxation policies and structures related to personal income tax rates implemented by different countries in the specified year.

`data/PSPR_incidence_dirtax_2023.dta`

The dataset utilized in the World Bank Poverty and Shared Prosperity report Chapter 5 "Correcting Course" Bank (2022) is sourced from the Commitment for Equity Institute and the World Bank. It was shared by the World Bank in response to the authors' request (Bank 2022). The dataset encompasses various variables, including country and year of the CEQ project (ctry_year), country of the CEQ project (ctry_ceq), year of the CEQ project (year), the incidence of direct taxes across income deciles (in_pdi_dirtax), World Economic Outlook group classification (class_weo), and others (Bank 2022). These variables show poverty, inequality, and shared prosperity metrics, enabling detailed analyses of economic and equity-related indicators across different countries and periods.

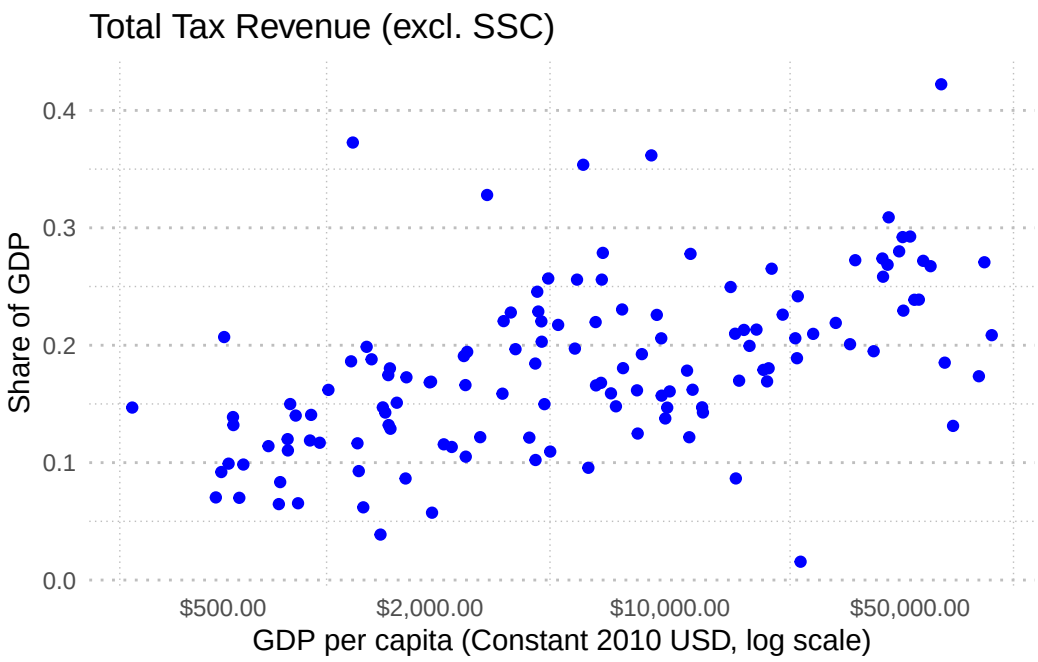

Figure 1: Relationship between tax revenue and gdp per capita

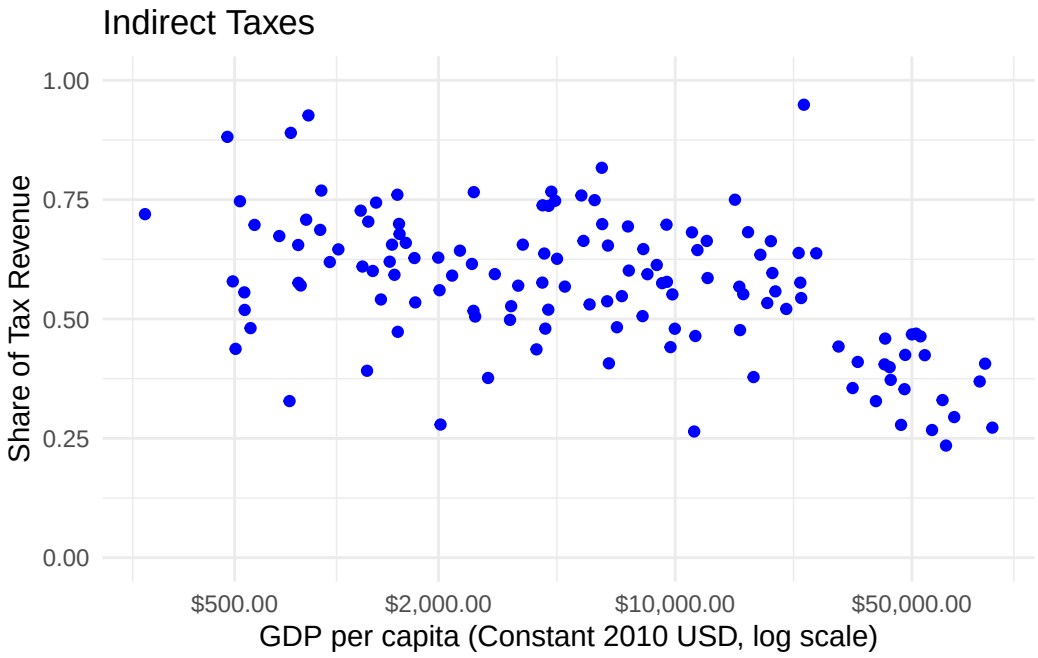

Figure 2: Relationship between tax revenue and gdp per capita

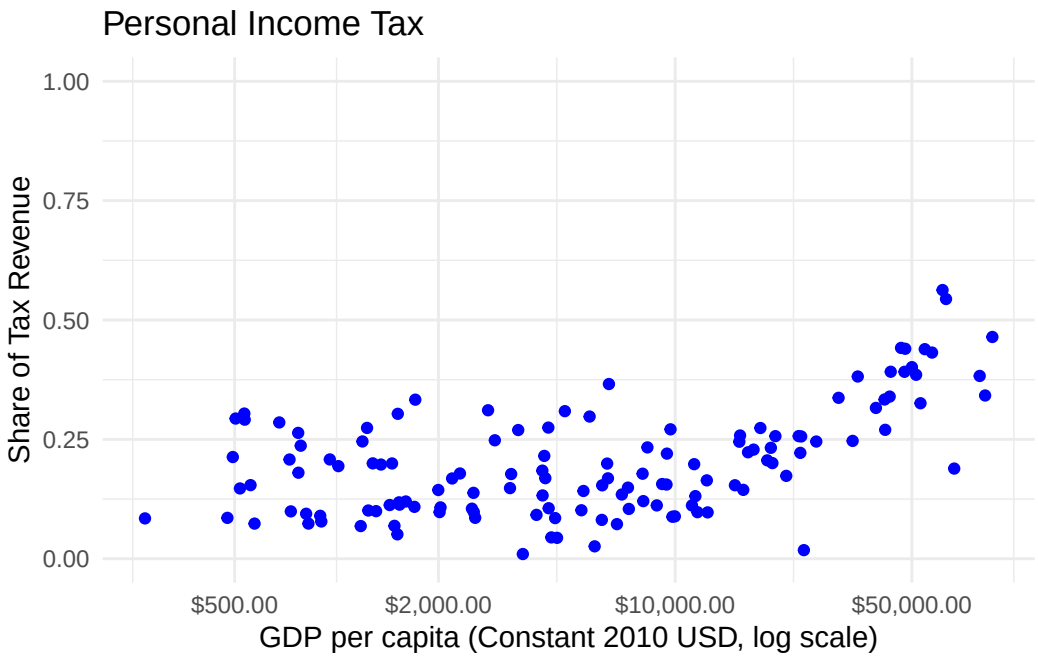

Figure 3: Relationship between tax revenue and gdp per capita

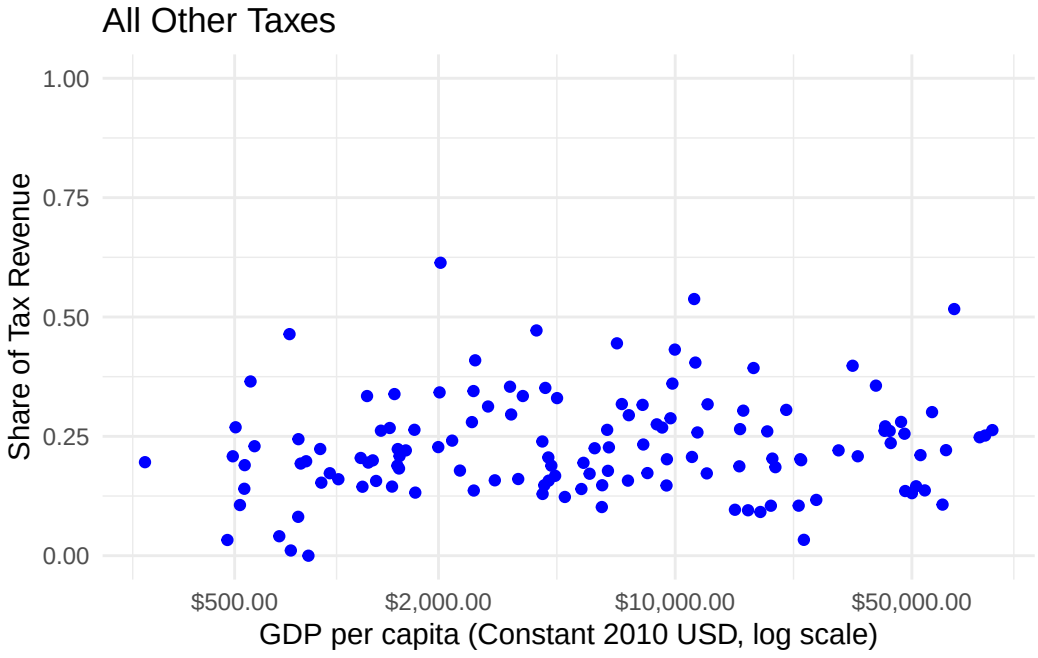

Figure 4: Relationship between tax revenue and gdp per capita

Figure 5: Relationship between the difficulties of collecting taxes and gdp per capita

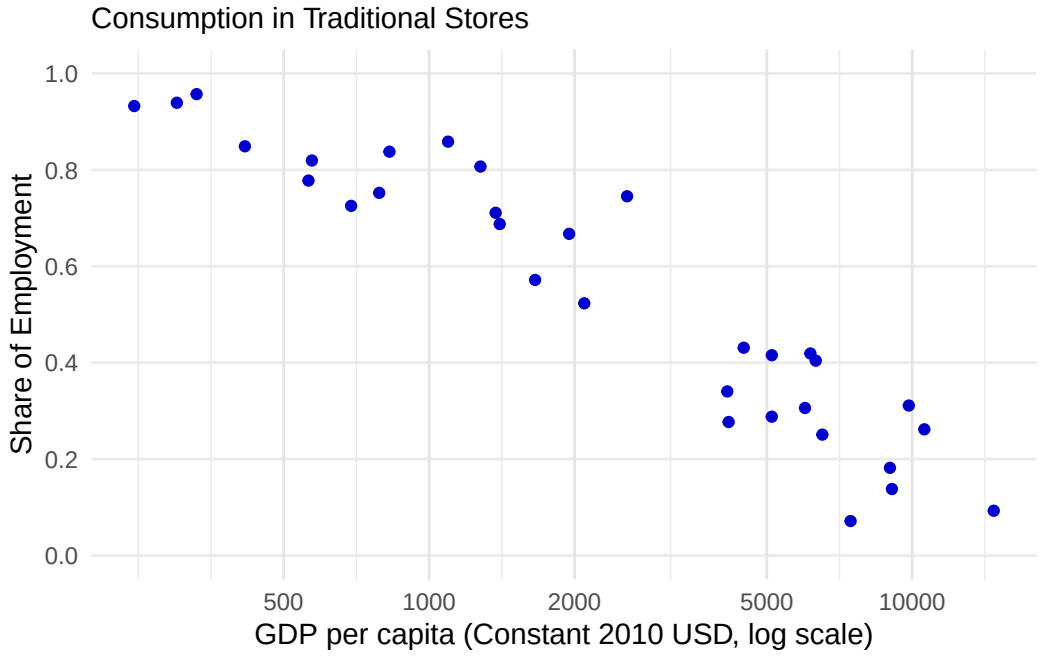

Figure 6: Relationship between the difficulties of collecting taxes and gdp per capita

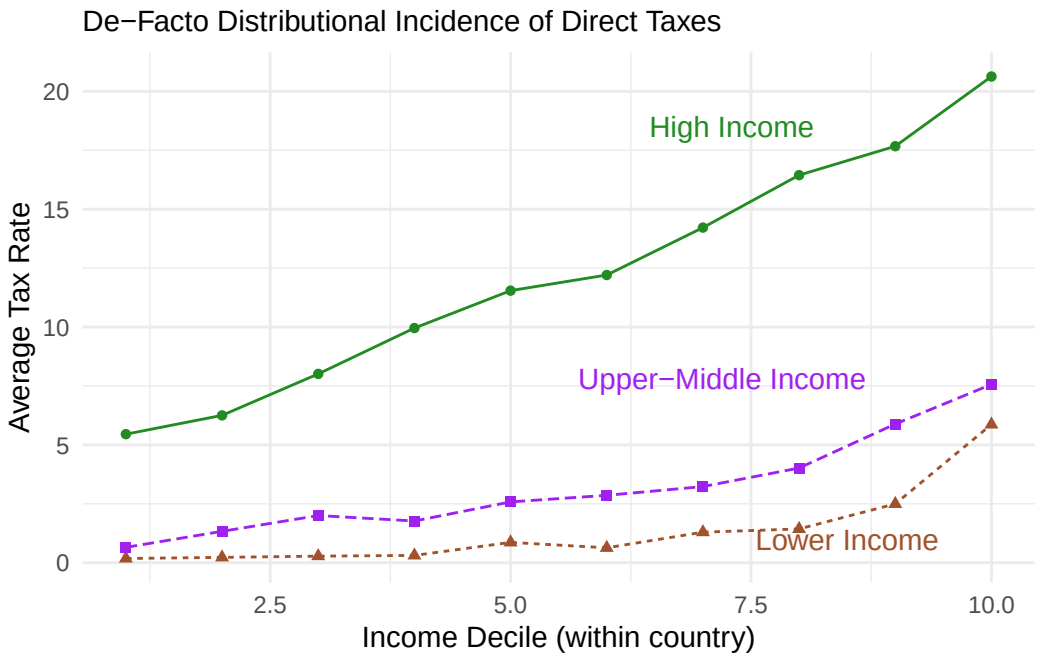

Figure 7: Relationship between the tax rate and gdp per capita

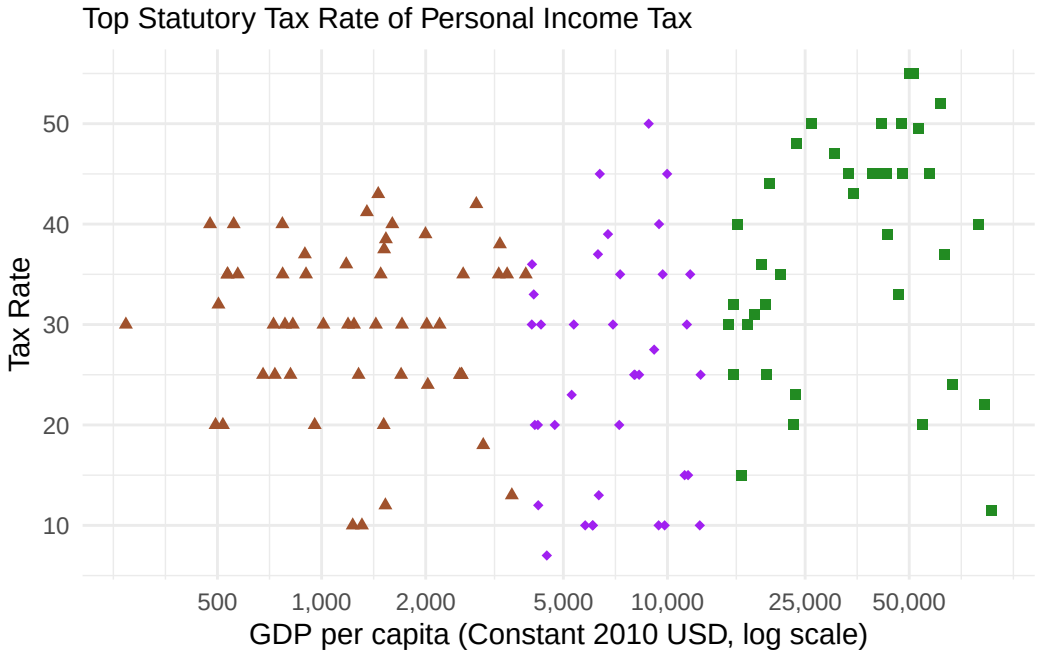

Figure 8: Relationship between the tax rate and gdp per capita

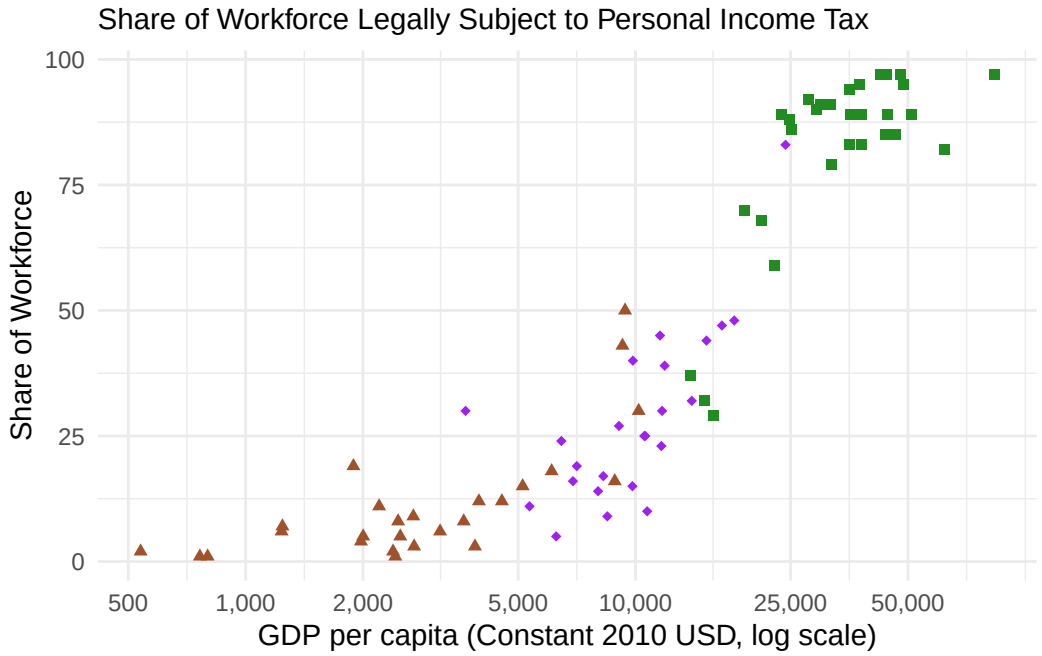

Figure 9: Relationship between the tax rate and gdp per capita

3 Results

3.1 Figure 1

3.2 Figure 2

3.3 Figure 4

4 Discussion

4.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

4.2 Second discussion point

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

References

Bachas, Pierre, Matthew H Fisher-Post, Anders Jensen, and Gabriel Zucman. 2022a. "Globalization and Factor Income Taxation." Working Paper 29819. Working Paper Series. National Bureau of Economic Research. https://doi.org/10.3386/w29819.

———. 2022b. "Globalization and Factor Income Taxation." *Globalization and Factor Income Taxation*. https://globaltaxation.world/.

Bachas, Pierre, Lucie Gadenne, and Anders Jensen. 2023. "Informality, Consumption Taxes, and Redistribution." *The Review of Economic Studies*, September. https://doi.org/10.1093/restud/rdad095.

Bachas, Pierre, Anders Jensen, and Lucie Gadenne. 2024. "Tax Equity in Low- and Middle-Income Countries." *Journal of Economic Perspectives* 38 (1): 55–80. https://doi.org/10.1257/jep.38.1.55.

Bank, World. 2018. "World Development Indicators." *The World Bank*. https://databank.worldbank.org/source/world-development-indicators.

———. 2022. "Poverty and Shared Prosperity 2022." *The World Bank*. https://www.worldbank.org/en/publication/poverty-and-shared-prosperity.

———. 2023. "World Bank Country and Lending Groups." *The World Bank*. https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groupscountry_frame.dta.

ILOSTAT. 2021. "Data Tools to Find and Download Labour Statistics." *International Labor Organization*. https://ilostat.ilo.org/data/.

Jensen, Anders. 2022. "Employment Structure and the Rise of the Modern Tax System." *American Economic Review* 112 (1): 213–34. https://doi.org/10.1257/aer.20191528.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ross, Michael, and Paasha Mahdavi. 2015. "Oil and Gas Data, 1932-2014." Harvard Dataverse. https://doi.org/10.7910/DVN/ZTPW0Y.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019a. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

———, et al. 2019b. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, Kirill Müller, and Davis Vaughan. 2023. *Dplyr: A Grammar of Data Manipulation*. https://dplyr.tidyverse.org.

Wickham, Hadley, Jim Hester, and Jennifer Bryan. 2024. *Readr: Read Rectangular Text Data*. https://readr.tidyverse.org.

Wickham, Hadley, Davis Vaughan, and Maximilian Girlich. 2024. *Tidyr: Tidy Messy Data*. https://tidyr.tidyverse.org.