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

Amie Liu, Wanling Ma

February 18, 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.

In this paper, we discuss how economic development influences tax structures, with a focus on the shift from indirect to personal income taxes as countries' economies grow, driven by an expanding formal sector and enhanced tax enforcement. Additionally, we examine how the prevalence of the informal sector affects tax progressivity, highlighting that indirect taxes like

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

VAT can become progressive in environments with significant informality. Lastly, we explore the challenges faced in taxing high-income individuals in low- and middle-income countries, including evasion and enforcement limitations, which impede the effectiveness of direct taxes on the wealthy.

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_population_WDI.dta`

The dataset was obtained from the World Bank's World Development Indicators World Bank (2018). It encompasses a comprehensive sample from 266 countries and spans from the year 1960 to 2020 (World 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 (World 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 (World 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 World 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) (World Bank 2023). The income level classification consists of five distinct variables: high income, low income, upper middle income, lower middle income, and unclassified (World 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_information.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" World 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 (World 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 (World 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.

3 Results

Figure 1 visualizes the relationship between a country's level of economic development, as denoted by GDP per capita (constant 2010 US dollars, on a logarithmic scale with breaks at $500, $2000, $10,000, and $50,000), and the total share of GDP allocated to tax revenue, excluding social security contributions (SSC). This exclusion targets a focus on tax revenues

more directly amenable to policy adjustments in the sphere of redistribution. The y-axis scale ranges from 0.0 to 0.5, indicating a span from no tax revenue to half of the GDP attributed to taxes.

Comprising data from 132 countries that satisfy the inclusion criteria of having a population above one million and deriving less than one-third of their GDP from oil and gas revenue, Figure 1 broadly demonstrates a pattern where the share of GDP collected as tax revenue increases concomitantly with GDP per capita. This observation is visualized in a trend line, modeled via a LOESS regression (with a 95% confidence interval shaded in grey), which gently ascends from left to right, suggesting a positive correlation between economic development (GDP per capita) and the proportion of GDP collected through taxes.

The majority of countries' data points cluster around the smoothed trend line, underscoring a general conformity to the depicted relationship. However, there are notable exceptions. A few outliers consist of countries with a relatively high GDP per capita (around $30,000) yet almost negligible tax revenue shares. The presence of outliers suggests that while a positive correlation exists, the relationship is not without its exceptions. This variance underlines the complexity of tax system dynamics and the factors influencing tax collection beyond mere economic development.

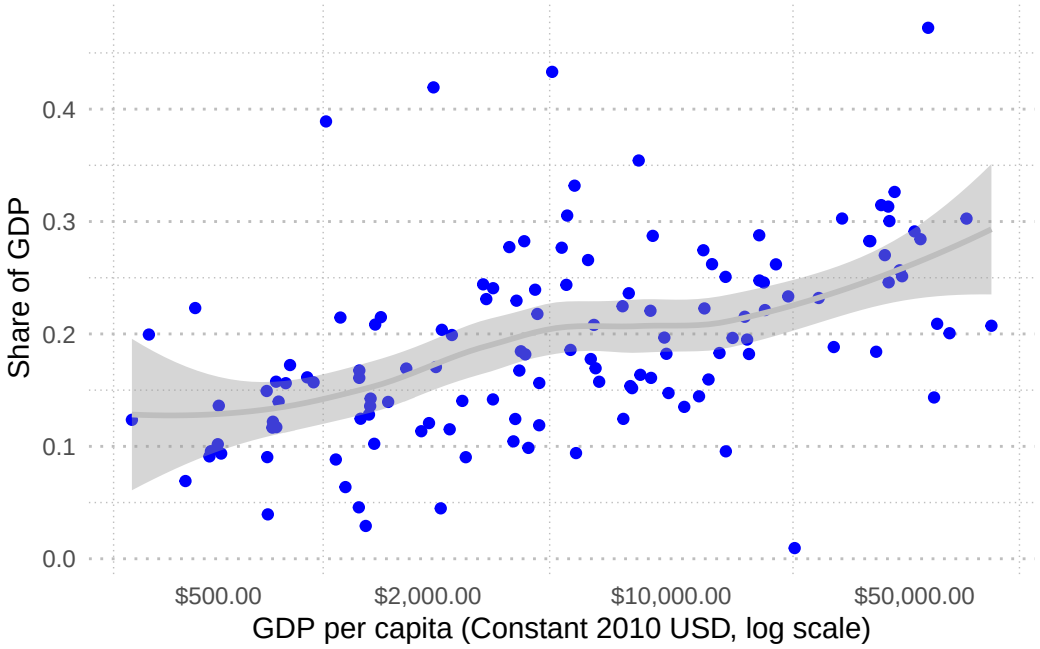

Figure 1: Total Tax Revenue (excl. SSC)

In the first graph of Figure 2, titled "Indirect Tax", the analysis reveals a gradually declining trend in the share of tax revenue derived from indirect taxes, such as the value-added tax, sales taxes, excises, and tariffs, as a country's GDP per capita increases. This depiction

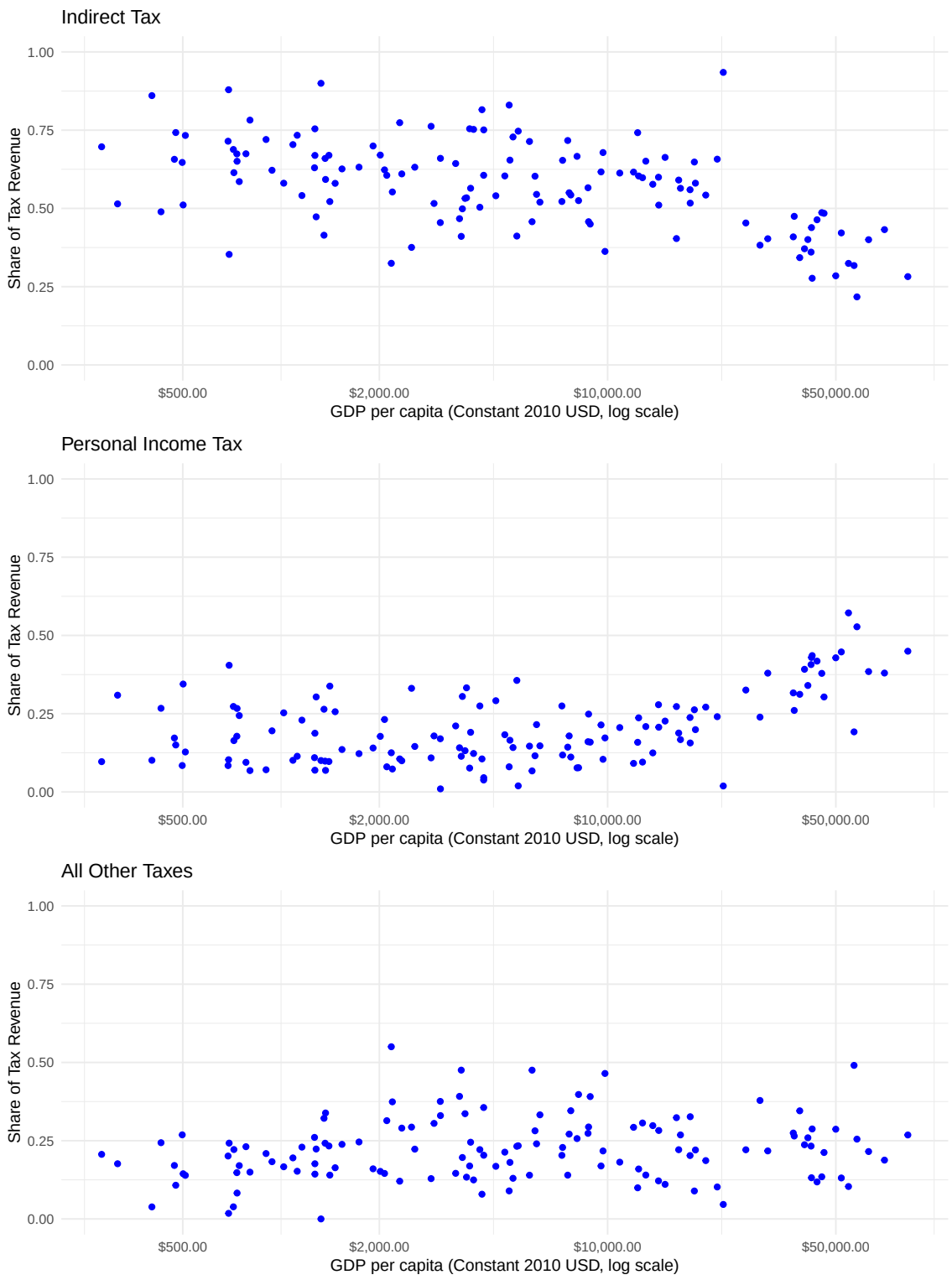

Figure 2: Share of Tax Revenue from different sources

underscores a noteworthy pattern wherein the reliance on indirect taxation decreases with economic advancement. The x-axis, representing GDP per capita, maintains the same log scale as in Figure 1, spanning from $500 to $50,000, while the y-axis scales from 0 to 1, indicating the proportion of total tax revenue attributed to indirect taxes. The uniformly gentle downward trajectory highlights a shifting tax structure in which nations with a lower GDP per capita have a higher propensity to lean on indirect taxes, a practice that progressively diminishes in favor of other tax forms as economies grow and develop.

Conversely, the second graph of Figure 2, showcasing "Personal Income Tax", presents a slight but consistently upward trend, suggesting that the share of personal income tax in total tax revenue tends to increase with a country's economic development. This pattern aligns with the transition observed in more developed economies, where direct taxation, particularly on personal income, becomes a more significant contributor to the tax base, reflecting a move towards more progressive taxation policies. In contrast, the third graph of Figure 2, "All Other Tax," exhibits a flat trend, indicating that the aggregate contribution of corporate income taxes, property, and wealth taxes to total tax revenue remains relatively stable across different levels of GDP per capita. Without the inclusion of confidence intervals in these panels, the clean, unobstructed trends in each graph offer a clear perspective on how the composition of tax revenue evolves with economic maturity, reinforcing the dynamic nature of tax systems and their responsiveness to broader economic contexts.

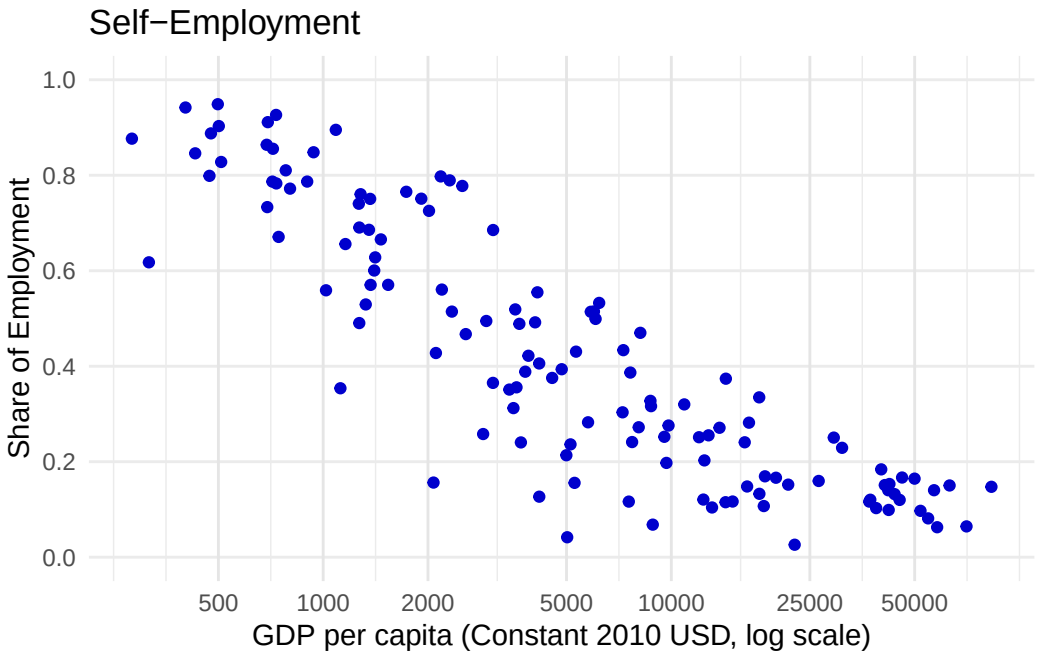

Figure 3: Self Employment

Figure 3 and Figure 4 articulately illustrates the inverse relationship between economic de-

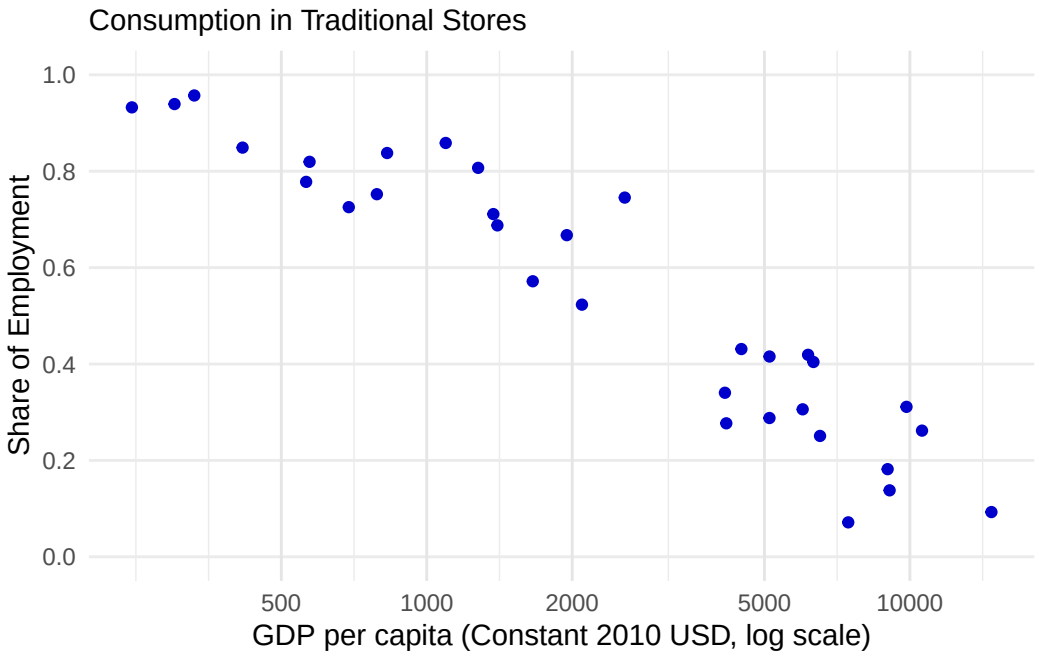

Figure 4: Consumption in Traditional Stores

velopment and the prevalence of informal economic activities, utilizing GDP per capita as a consistent measure of economic development across both figures. In Figure 3, a markedly steep downward trend is shown. The graph delineates the share of self-employment within the active workforce—a proxy indicator for informality in labor—plummeting as the GDP per capita increases. With the x-axis mirroring the range of GDP per capita found in Figure 1, and the y-axis proportionately scaling from 0 to 1 to represent the share of employment, the visualization clearly communicates a significant decrease in self-employment ratios, moving from low-income to high-income countries. This correlation underscores the observation that the formalization of labor concomitantly rises with economic prosperity, highlighting a transition from self-employed to formally employed work statuses as nations advance economically.

Similarly, Figure 4 exhibits an equally steep downward trajectory, mapping the share of total consumption allocated towards traditional retail outlets—such as street stalls, public markets, and corner stores—against GDP per capita. This axis configuration adheres to the same range as Figure 1, facilitating a direct comparison of informality trends across different dimensions of the economy. The decline from almost universal reliance on informal retail in low-income countries, denoted by a share approaching 1, to a minimal percentage in high-income countries, reaching down to 0.12, starkly exemplifies the diminishing role of informal consumption channels with advancing economic maturity. The parallel between the steep decline in both informal labor and consumption vividly encapsulates the broader economic transition from informality to formality, reinforcing the theme that economic growth aligns with a decreasing

share of the informal sector within the economy.

Figure 5 provides insights into how the average effective income tax rates vary across income deciles within 74 countries of varying income levels in 2022. These countries were selected based on the availability of fiscal incidence reports, with 30 categorized as high-income countries, 21 as upper-middle-income countries, and 23 as lower-income countries based on the World Bank (Bachas, Jensen, and Gadenne 2024). It highlights two main findings, which are higher average tax rates in higher income countries and the progressivity of income tax rates. We see that the average tax rates for income taxes are higher in high-income countries compared to low- and middle-income countries across all deciles. Specifically, the average tax rate for the highest income level (10th decile) in high-income countries exceeds 20%, while it remains below 10% for both upper-middle and lower-income countries (see Figure 5). Similarly, the average tax rate for the lowest income level (1st decile) surpasses 5% in high-income countries but remains below 5% in lower-income and middle-income countries.

Additionally, the figure illustrates the progressive nature of income tax rates, particularly in high-income countries. Within these countries, the average tax rates increase significantly across income deciles, from the 1st decile to the 10th decile of income distribution. While the households in the 10th decile in high-income countries bear more than 20% tax burden, the corresponding figures for low- and middle-income countries are 6% and 8% respectively (see Figure 5). This observation indicates the progressive structure of the income tax system, where households with higher incomes contribute a larger proportion of their income in taxes compared to those with lower incomes. These findings underscore the differences in average income tax rates between different income levels and deciles, which clarifies the progressive nature of income tax systems in high-income countries and emphasizes the importance of considering the level of economic development in the design of tax policies.

Figure 6 shows the top statutory tax rate of the personal income tax across 129 countries, utilizing data from 2023 collected by the authors (Bachas, Jensen, and Gadenne 2024) and particularly distinguishing between high-income countries and low- and middle-income countries. On average, the top statutory tax rate in high-income countries is marginally higher, standing at 35%, compared to 29% in low- and middle-income countries (see Figure 6). Despite the slight variation in top statutory tax rates, Figure 8 does not reflect a proportional relationship between GDP per capita and the top statutory tax rate. This underscores the nuanced dynamics of personal income taxation across different income levels and economic contexts.

Figure 7 shows the share of the active workforce subject to personal income tax across varying levels of economic development, as indicated by GDP per capita. It encompasses 92 countries and underscores the evolution of personal income taxation policies from lower-income to high-income countries (Bachas, Jensen, and Gadenne 2024). It reveals a distinct trend: as economic development progresses, there is a gradual increase in the share of the workforce subject to personal income tax obligations (see Figure 7). In lower-income countries, the coverage of the personal income tax is minimal, with the lowest point hovering slightly above 0 share of the

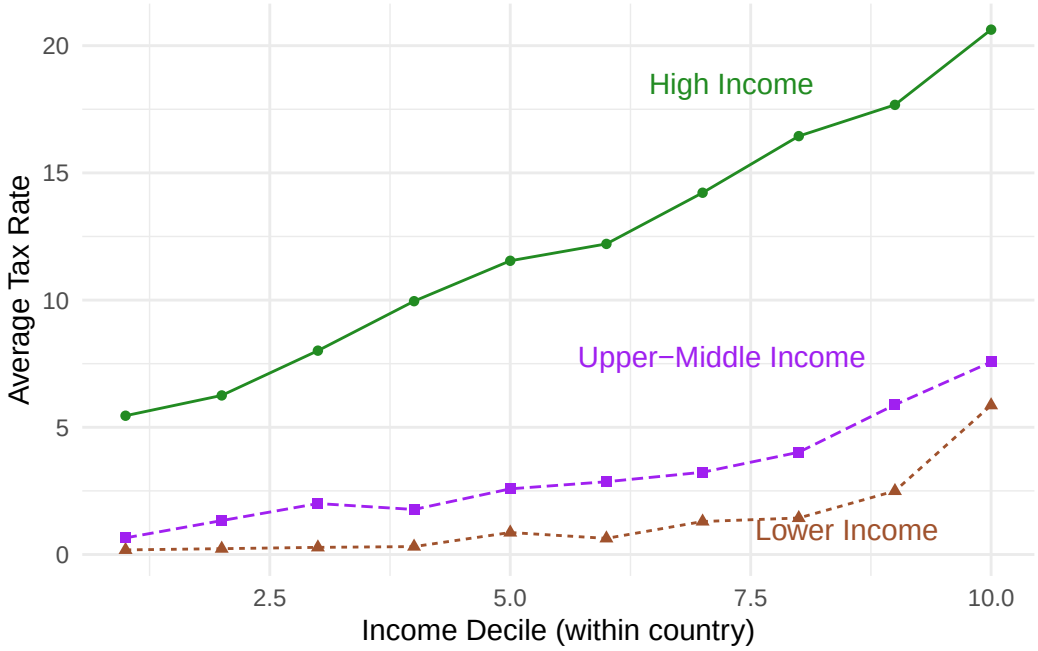

Figure 5: De-Facto Distributional Incidence of Direct Taxes

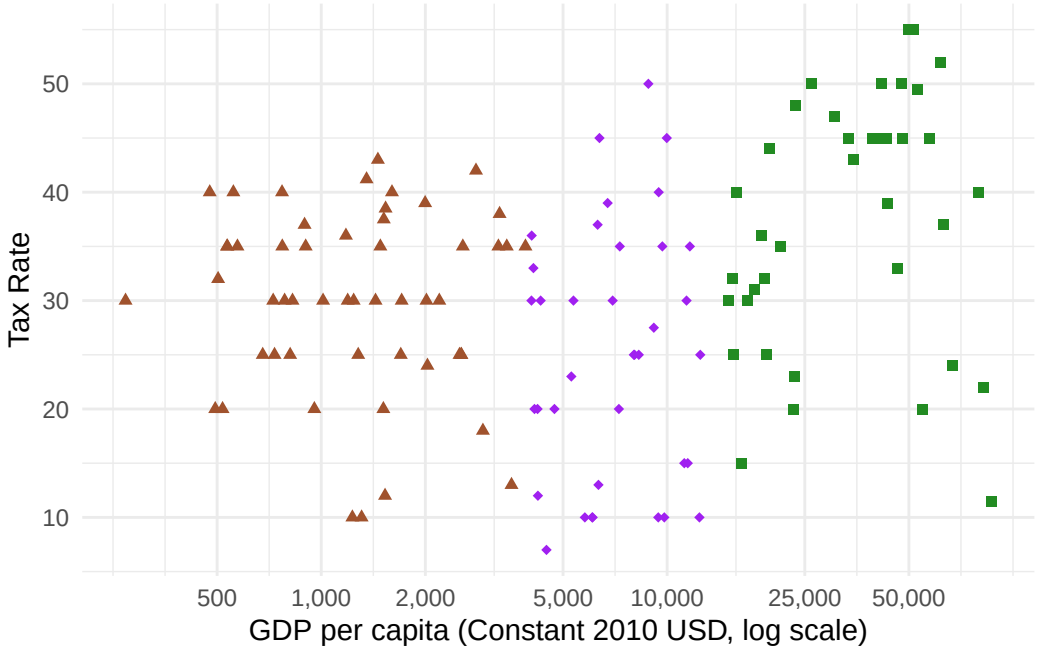

Figure 6: Top Statutory Tax Rate of Personal Income Tax

workforce and slightly over 500 GDP per capita. Conversely, high-income countries demonstrate a significantly higher share of the workforce population legally liable to pay personal income taxes, with the highest point approaching nearly 100 share of the workforce and exceeding 50,000 GDP per capita (see Figure 7). These key findings highlight the differences in the implementation and coverage of personal income taxation policies between high-income and low- and middle-income countries, showing the evolution of tax systems across varying levels of economic development.

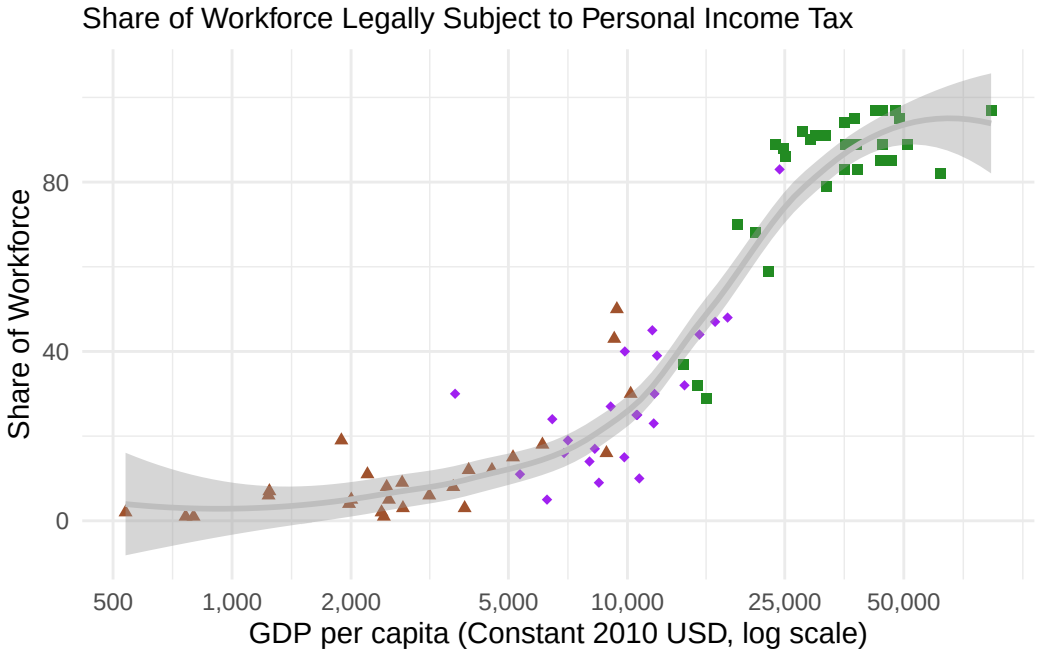

Figure 7: Share of Workforce Legally Subject to Personal Income Tax

4 Discussion

4.1 Tax Structure and Economic Development

4.2 Informality and Taxation

4.3 Ethics and Biases

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.

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.

World Bank. 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.