

A New Database of General Government Revenue and Expenditure by Function, 1980-2022

Amory Gethin*

March 1, 2024

Abstract

This technical note presents new series of general government expenditure in all countries in the world since 1980. Combining data from eighteen sources, I estimate the level and composition of public spending by function of government, with a particular focus on education, healthcare, and social protection. I also compile estimates of general government revenue and its composition, as well as series of debt service cost and government primary budget balance. Despite severe data limitations in some countries, the resulting database opens new avenues for the study of the evolution of the size of governments and the role of public spending in reducing poverty and inequality.

*Paris School of Economics (Grant ANR-17-EURE-001).

1. Introduction

This technical note presents the methodology used to estimate a new database on the level and composition of general government expenditure in all countries in the world from 1980 to 2022. This database serves as the basis for the analysis of the role of public services in global poverty reduction developed in [Gethin \(2023\)](#), as well as for the construction of comparable estimates of posttax income inequality developed in [Fisher-Post and Gethin \(2023\)](#).

Section 2 presents data sources. Section 3 presents the methodology used to harmonize total general government expenditure series. Sections 4, 5, 6, 7, and 8 turn to the harmonization of education, health, social protection, other components of government expenditure, and debt service cost, respectively. Section 9 covers general government revenue.

2. Data Sources

To construct comparable estimates of general government revenue and expenditure, I collect data from a number of available sources.

Data on general government expenditure and its composition by function of government come from eight sources:

1. IMF historical government revenue and expenditure series, based on initial data compilation by [Mauro et al. \(2015\)](#) and updates.¹ The data cover total government revenue, total government expenditure, interest paid on public debt, and government primary balance for over 150 countries from 1800 to 2022 (unbalanced panel).
2. IMF series of government expenditure by function.² The data cover total expenditure and its breakdown by function (COFOG) for different levels of government, including general government, central government, state government, and local government series.
3. IMF series of government expenditure by function from historical government finance statistics. This data source is comparable to the previous one but covers earlier years.³
4. Eurostat series of government expenditure by function, covering the European Union.⁴

¹<https://www.imf.org/external/datamapper/datasets/FPP>

²<https://data.imf.org/?sk=388dfa60-1d26-4ade-b505-a05a558d9a42&sid=1479329334655>

³<https://www.icpsr.umich.edu/web/ICPSR/studies/8624/publications>

⁴<https://ec.europa.eu/eurostat/data/database>

5. OECD series of government expenditure by function, covering OECD countries.⁵
6. CEPAL series of government expenditure by function, covering Latin American countries.⁶
7. The SPEED database (Statistics on public expenditures for economic development) maintained by the International Food Policy Research Institute (IFPRI), compiling series of government expenditure by function from various sources.⁷
8. United Nations series of government final consumption expenditure by function.⁸

Additional data on education and health expenditure come from:

1. The UNESCO UIS statistics, covering series of public education expenditure by level (primary, secondary, tertiary).⁹
2. The UNESCO's World Education Report 1991, which provides additional data points for 1980 and 1988 in a number of countries.
3. [Bharti and Yang \(2023\)](#), covering China and India since 1980.
4. The World Health Organization, covering health expenditure in almost all countries in the world since 2000 (available from the World Bank's World Development Indicators).

Additional data on social protection expenditure come from:

1. The Asian Development Bank, covering total social protection expenditure in a number of Asian countries.¹⁰
2. The CEPAL SOCX database, covering social protection expenditure and its decomposition by type of program in Latin American countries.¹¹
3. The OECD SOCX database, covering social protection expenditure and its decomposition by type of program in OECD countries.¹²

⁵https://stats.oecd.org/Index.aspx?DataSetCode=SNA_TABLE11

⁶<https://statistics.cepal.org>

⁷<https://www.ifpri.org/publication/statistics-public-expenditures-economic-development-speed>

⁸<http://data.un.org/Data.aspx>

⁹<http://data.uis.unesco.org/>

¹⁰<https://spi.adb.org/spidmz/>

¹¹<https://statistics.cepal.org>

¹²https://stats.oecd.org/Index.aspx?datasetcode=SOCX_AGG

4. The World Bank SPEED database, covering social protection expenditure and its decomposition by type of program in selected Eastern European and Central Asian countries.¹³
5. The World Bank ASPIRE database, covering social protection expenditure and its decomposition by type of program in a number of developing countries.¹⁴

3. General Government Expenditure

I start by constructing harmonized series of total general government expenditure for all countries. I do so in two steps.

First, I attribute a “benchmark series” to each country, corresponding to the primary data source used to observed the level of total expenditure. I assign one benchmark source to each country, with the following order of priority: OECD, CEPAL, IMF historical series, other IMF series (general government when available, or the sum of central and local governments, or central government), SPEED, and the World Bank. For countries not covered by any of these sources, I assume that expenditure equals total general government tax revenue, as reported in [Bachas et al. \(2022\)](#). For countries still not covered, I use regional averages of expenditure as a share of GDP.

Second, I use other sources to extend benchmark series backwards and forwards. For instance, if the OECD covers the 2000-2022 period in a given country while the IMF covers 1980-2022, I extrapolate OECD series backwards using absolute changes in expenditure as a percentage of GDP available from the IMF over the 1980-2000 period. This ensures that harmonized series do not have sudden ups and downs due to inconsistencies across sources, while taking the best of available information on changes in expenditure over time from each source. I follow the same order of priority as above. As a last resort, to go back to 1980, I assume that total expenditure has remained constant as a share of GDP when no other information is available.

4. Education

Given that education spending series have greater coverage than series providing the full breakdown of expenditure by function of government, I choose to separately estimate education spending rather than directly estimating the full decomposition of government expenditure. The same principle holds for healthcare and social protection. I thus construct series of education,

¹³<https://www.worldbank.org/en/programs/speed>

¹⁴<https://www.worldbank.org/en/data/datatopics/aspire/indicator-glance>

health, and social protection spending separately, and then combine them with other information on the composition of expenditure to estimate other functions of government as a residual.

To estimate harmonized public education expenditure series, I follow the same principles as those used for total expenditure.

First, I attribute a benchmark series by order of priority: general government education expenditure from Eurostat, the OECD, CEPAL, the IMF, and the UNESCO, followed by central government expenditure from the same sources if no general government expenditure series is available. Data for China and India are taken from [Bharti and Yang \(2023\)](#).

Second, I carry these benchmark series backward and forward using other available sources, with the same order or priority. To cover the full 1980-2022 period, I assume that education expenditure has remained constant as a share of total general government primary expenditure when no other information is available.

Third, I estimate the breakdown of education expenditure by level (primary, secondary, tertiary). I combine all available data points from UNESCO, CEPAL, and the IMF, by order of priority. Data for China and India are taken from [Bharti and Yang \(2023\)](#). To cover the full 1980-2022 period, I assume that the composition has remained constant when no other information is available.

5. Health

I follow similar steps to estimate harmonized series of public health expenditure. The benchmark series are general government health expenditure from Eurostat, the OECD, CEPAL, the IMF, the WHO, and SPEED, by order of priority. In addition to WHO series available online, which cover the 2000-2022 period, I add historical series from [Murray, Govindaraj, and Musgrove \(1994\)](#), which report estimates of public health expenditure in 153 countries in 1990. I carry these series backward and forward using sources other than the benchmark, as for education. To cover the full 1980-2022 period, I assume that health expenditure has remained constant as a share of total general government primary expenditure when no other information is available.

6. Social Protection

I follow similar steps to estimate harmonized series of social protection expenditure.

First, I construct series of total social protection expenditure, including social insurance and social assistance programs. The benchmark series are general government social protection

expenditure from Eurostat, the OECD, CEPAL, and the IMF, followed by central government expenditure from the same sources if no general government expenditure series is available. For countries with no data from any of these sources, I use estimates from the World Bank's ASPIRE database, the United Nations' final consumption expenditure series, the SPEED database, or the Asian Development Bank, depending on which of these sources is available and seems most consistent. I carry these series backward and forward using sources other than the benchmark, as for education and health. To cover the full 1980-2022 period, I assume that social protection expenditure has remained constant as a share of total general government primary expenditure when no other information is available.

Second, I split social protection expenditure into social insurance and social assistance. For consistency, I only use one source by country: the OECD SOCX database, the CEPAL SOCX database, and the World Bank ASPIRE database, by order of priority. For the 41 countries with data from none of these sources, I use regional averages across countries with available data. To cover the full 1980-2022 period, I assume that this share has remained constant when no data is available.

Third, I split social assistance expenditure into cash and in-kind transfers. This breakdown is available from the OECD SOCX database and the World Bank ASPIRE database, covering together 127 countries. In the remaining countries, I use regional averages across countries with available data. To cover the full 1980-2022 period, I assume that this share has remained constant when no data is available.

7. Other Components of Government Primary Expenditure

The other components of government primary expenditure are expenditure on general public services (excluding debt service cost), defense, public order and safety, economic affairs, environment protection, housing and community amenities, and recreation, culture and religion. Data on these components of expenditure are much scarcer and much more difficult to link across sources. I estimate them in two steps.

First, I construct series on the composition of expenditure covering all functions of government. Given difficulties in linking series, I assign only one source by country: general government expenditure series from Eurostat, the OECD, CEPAL, and the IMF, by order of priority, followed by central government expenditure series (or the sum of local and central governments) from the same sources when general government series are not available. For countries with no data from any of these sources, I use regional averages of the composition of spending by COFOG.

Second, I ensure consistency between these series and the total expenditure, health expenditure, education expenditure, and social protection expenditure series estimated separately above. I assume that these four series are correct and proportionally adjust the relative importance of other functions of government. This amounts to estimating the size of these other functions of government as a residual (total expenditure minus education, health, and social protection), while using the above sources to measure their importance relative to one another. In the rare cases where this leads to negative value (that is, the sum of education, health, and social protection exceeds total expenditure), I assume that the education, health, and social protection series are correct, that other functions of government represent 1% of total expenditure each, and I adjust total expenditure series upwards accordingly.

8. Debt Service Cost

I estimate debt service cost series separately, using the same methodology as for education, health, and social protection expenditure. The benchmark series are historical IMF series, general government expenditure IMF series, and central government expenditure IMF series, by order of priority. To cover the full 1980-2022 period, I assume that debt service cost has remained constant as a share of total general government expenditure when no other information is available. For missing countries, I use regional averages of debt service cost as a share of GDP. Having estimated debt service cost, I calculate general government primary expenditure, equal to total general government expenditure minus debt service cost.

9. General Government Revenue

Finally, I estimate the level and composition of general government revenue using the same methodology. The benchmark series are historical IMF series, which I carry backward and forward using total tax revenue estimates from [Bachas et al. \(2022\)](#). In the few cases where total tax revenue from [Bachas et al. \(2022\)](#) exceeds IMF series, I assume that total government revenue equals tax revenue. I then calculate non-tax revenue as the difference between general government revenue and tax revenue. The composition of tax revenue by type of tax directly comes from [Bachas et al. \(2022\)](#), including personal income taxes, corporate income taxes, property and wealth taxes, indirect taxes, other taxes, and social contributions. To cover the full 1980-2022 period, I assume that government revenue as a share of GDP and its composition have remained constant when no data is available. In missing countries, I use regional averages

of government revenue as a share of GDP

References

- Bachas, Pierre, Matthew H. Fisher-Post, Anders Jensen, and Gabriel Zucman (2022). “Globalization and Factor Income Taxation”. NBER Working Paper 29819.
- Bharti, Nitin and Li Yang (2023). “Human Capital Accumulation in China and India in 20th century”. Working Paper.
- Fisher-Post, Matthew and Amory Gethin (2023). “Government Redistribution and Development: Global Estimates of Tax-and-Transfer Progressivity, 1980-2019”. Working Paper.
- Gethin, Amory (2023). “Revisiting Global Poverty Reduction: Public Goods and the World Distribution of Income, 1980-2019”. Working Paper: <https://amory-gethin.fr/files/pdf/Gethin2023PublicGoods.pdf>.
- Mauro, Paolo, Rafael Romeu, Ariel Binder, and Asad Zaman (2015). “A modern history of fiscal prudence and profligacy”. In: *Journal of Monetary Economics* 76.1, pp. 55–70.
- Murray, C. J. L., R. Govindaraj, and P Musgrove (1994). “National health expenditures: a global analysis”. In: *Bull World Health Organ* 72.4, pp. 623–637.