

Global Macro Database

Technical Appendix

This appendix contains additional technical details on the construction of the time series introduced in the paper *The Global Macro Database: A New International Macroeconomic Dataset* by Karsten Müller, Chenzi Xu, Mohamed Lehib, and Ziliang Chen. Further detailed information specific to individual observations, countries, and variables can be found on www.globalmacrodata.com/documentation.

Table of Contents

TA.1 Acknowledgments	TA-4
TA.2 Data Coverage	TA-5
TA.2.1 Dataset overview	TA-6
TA.2.2 Comparison with other sources	TA-9
TA.3 Sources and Data Access	TA-14
TA.3.1 Types of data sources	TA-14
TA.3.2 How we update the data	TA-20
TA.3.3 Data storage and version control	TA-20
TA.3.4 Data update process	TA-21
TA.4 Variable Definitions	TA-21
TA.4.1 Statistical framework	TA-21
TA.4.2 Main national account variables	TA-22
TA.4.3 Consumption and investment	TA-22
TA.4.4 Investment	TA-23
TA.4.5 External sector	TA-23
TA.4.6 Government finance	TA-24
TA.4.7 Money and interest rates	TA-26
TA.4.8 Prices and labor market	TA-27
TA.5 Priority Ordering of Data Sources	TA-28
TA.5.1 First tier: Modern official sources	TA-28
TA.5.2 Second tier: Country-specific historical sources	TA-28
TA.5.3 Third tier: Other aggregators	TA-28
TA.6 Measurement Issues	TA-29
TA.6.1 Data quality	TA-29
TA.6.2 Dealing with data revisions	TA-30

TA.6.3	Base years for index variables	TA-32
TA.6.4	Changes in currency	TA-33
TA.6.5	Dealing with series breaks	TA-35
TA.6.6	Doubtful data points	TA-36

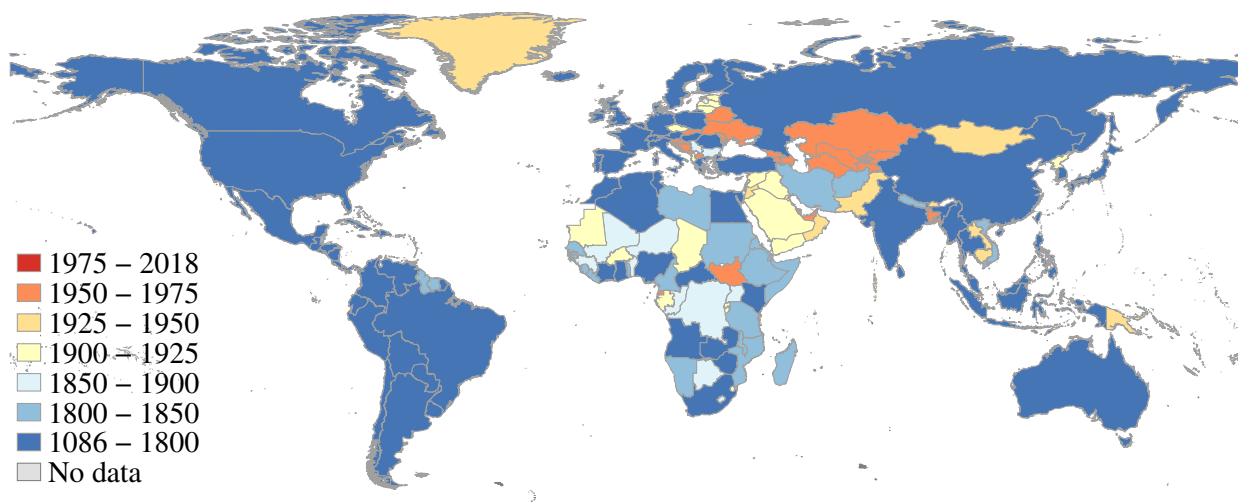
TA.1 Acknowledgments

The development of the Global Macro Database would not have been possible without the generous funding provided by the Singapore Ministry of Education (MOE) through the PYP grants (WBS A-0003319-01-00 and A-0003319-02-00), a Tier 1 grant (A-8001749-00-00), and the NUS Risk Management Institute (A-8002360-00-00). This financial support laid the foundation for the successful completion of this extensive project.

TA.2 Data Coverage

The Global Macro Database comprises a collection of 46 macroeconomic variables, spanning 243 countries and territories. The majority of countries have time series data extending to pre-1950 periods, with many reaching back to the early 20th century or even the 19th century, including several developing nations. Figure TA1 presents a world map indicating the earliest year for which any macroeconomic statistic is available for each country (excluding population data).

Figure TA1: First Year in Dataset, By Country



Note: This figure visualizes, for each country, the first year for which we have data on any macroeconomic variable other than population in the Global Macro Database (GMD). The map demonstrates that for the vast majority of countries, we have time series before 1950, with many series extending back to the 1800s for both advanced and developing economies. For some countries, such as the United Kingdom, data coverage begins as early as 1086.

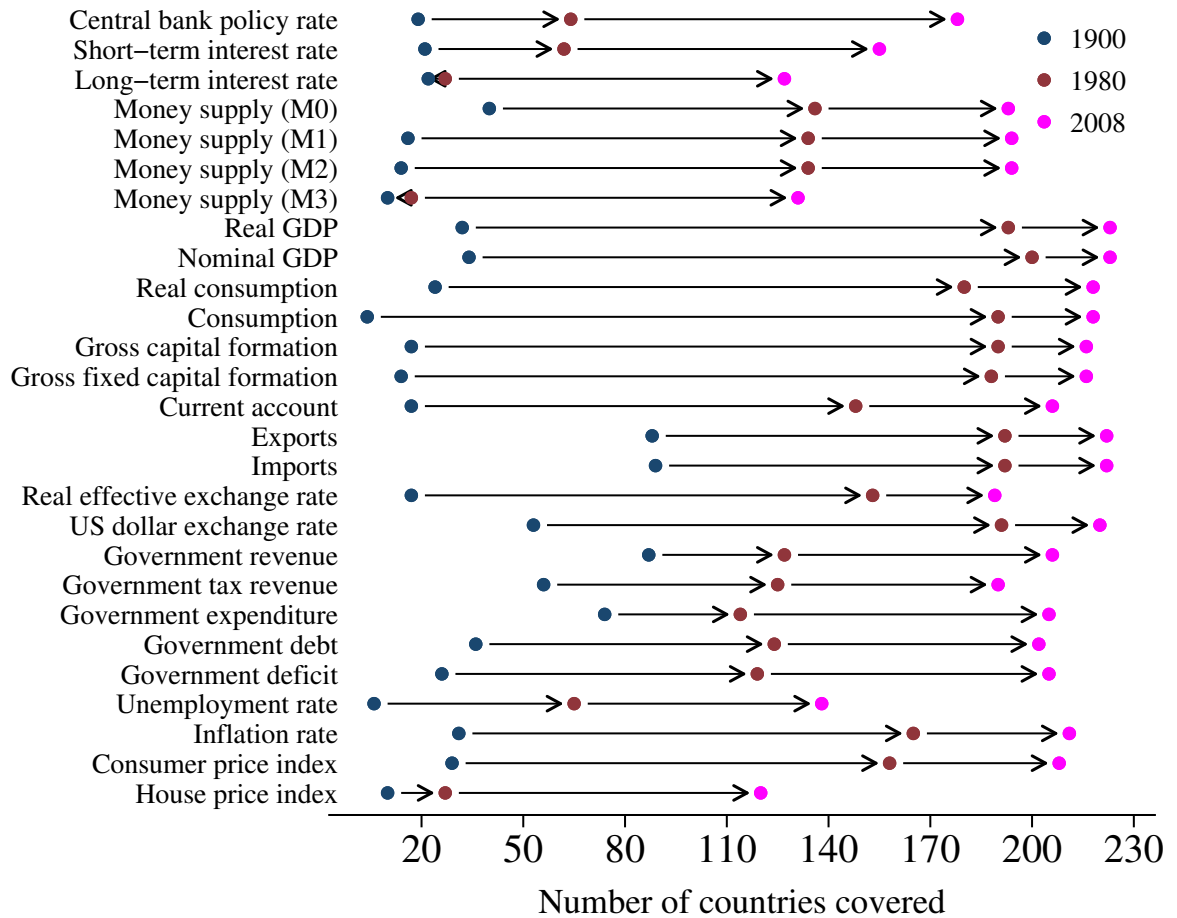
TA.2.1 Dataset overview

The following tables provide an overview of the key variables, along with their respective time and country coverage, highlighting the dataset's extensive scope and usefulness for cross-country and panel analyses.

Table [TA1](#) shows the key variables included in the Global Macro Database (GMD). The table provides detailed information on each variable, including the variable name, the abbreviation used in the database, units of measurement, temporal coverage, forecast horizons, and country coverage. The variables are grouped into six categories: national accounts, consumption and investments, external sectors, government finances, money and interest rates, and prices and labor market. For detailed definitions and descriptions of each variable, please refer to Section [TA.4](#).

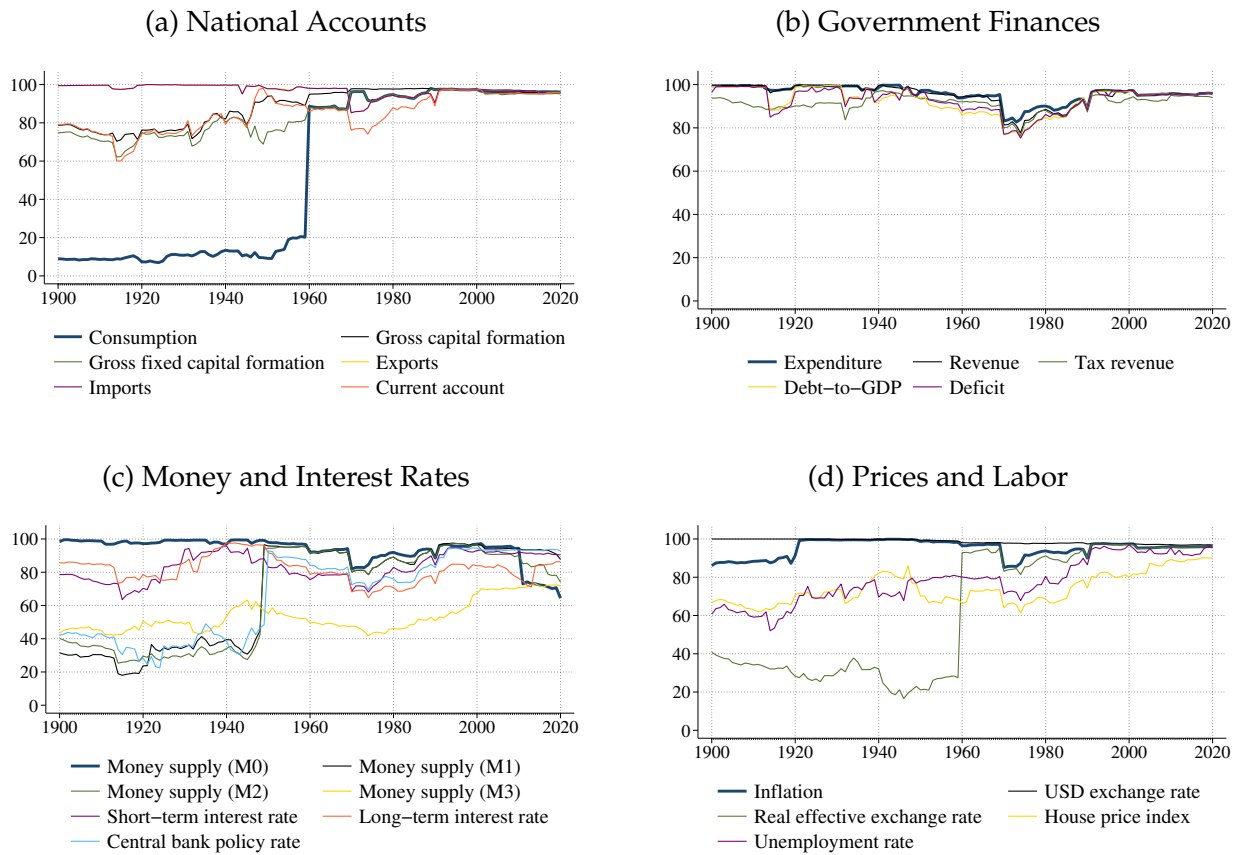
Figure [TA2](#) illustrates the number of countries covered by the Global Macro Database (GMD) for each variable in the years 1900, 1980, and 2008. A significant proportion of the variables have coverage for over 20 countries since 1900, over 100 countries since 1980, and nearly 200 countries since 2008, reflecting the increasing scope and comprehensiveness of the database over time. Figure [TA3](#) depicts the share of GDP for which data is available on key variables in the GMD.

Figure TA2: Number of Countries Covered, By Variable



Note: This figure plots the number of countries with available data for all macroeconomic variables contained in the Global Macro Database (GMD) at three points in time: 1900, 1980, and 2008. Each line represents a variable, with dots indicating the coverage at these three dates. The graph reveals substantial variation in coverage across variables and time.

Figure TA3: Share of GDP Covered Over Time, By Variable



Note: This figure plots the share of GDP covered by each variable in the Global Macro Database (GMD) between 1900 and 2020.

TA.2.2 Comparison with other sources

By design, the coverage of our dataset surpasses that of all existing publicly available sources, as these sources have been integrated into the Global Macro Database (GMD). Table [TA2](#) compares the coverage of key variables in the GMD with those offered by other widely-used data providers. Table [TA3](#) presents the number of country-year observations included in the GMD and the fraction covered by the major providers in percentage. Figure [TA4](#) compares the coverage of the Global Macro Database (GMD) with that of the next best source for each key variable.

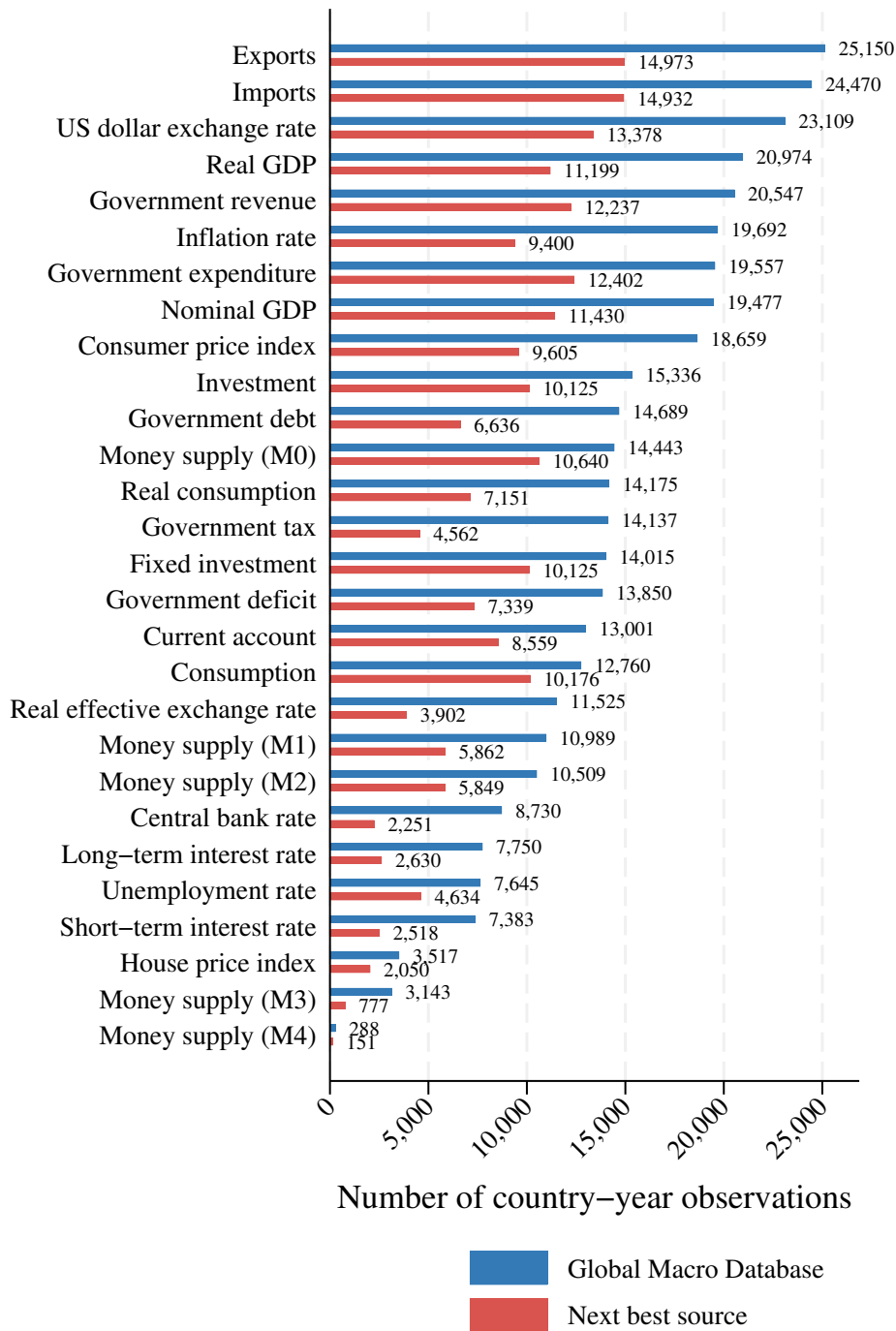
In addition to publicly available sources, Table [TA2](#) also includes a comparison with the dataset provided by the commercial data provider Global Financial Data (GFD). Beyond surpassing the coverage of many commercial products, the Global Macro Database (GMD) offers more comprehensive and transparent documentation, supported by an open-source code base. For further details, please refer to our [GitHub repository](#).

Table TA3: Coverage of GMD Variables in Selected Sources

Variable	GMD	Fraction of observations in the GMD covered in... (in %)							
		IFS*	WEO	OECD EO	WDI†	UN	JST	Mitchell	GFD‡
Central bank policy rate	8,730	26	—	21	—	—	—	—	73
Short-term interest rate	7,383	24	—	28	—	—	34	—	56
Long-term interest rate	7,750	10	—	23	—	—	34	—	100
Money supply (M0)	14,455	47	—	—	—	—	3	74	93
Money supply (M1)	10,989	61	—	11	—	—	18	53	42
Money supply (M2)	10,509	64	—	—	—	—	14	56	—
Money supply (M3)	3,143	—	—	31	—	—	25	—	—
Money supply (M4)	288	—	—	—	—	—	52	—	—
Real GDP	20,974	25	42	11	53	48	13	31	95
Nominal GDP	19,477	30	45	12	59	52	14	39	84
Consumption	12,752	24	—	18	66	80	—	—	—
Gross capital formation	15,412	27	50	15	54	66	16	36	28
Gross fixed capital formation	14,007	31	—	16	56	72	—	32	28
Current account	13,001	33	66	16	57	—	19	—	21
Exports	25,167	18	39	9	35	35	10	59	73
Imports	24,487	18	40	9	36	36	11	61	75
Real effective exchange rate	11,525	34	—	8	34	—	—	—	39
US dollar exchange rate	23,109	58	—	12	48	—	12	—	98
Government revenue	20,515	20	36	6	22	—	12	60	71
Government tax revenue	14,097	29	—	7	32	—	—	32	1
Government expenditure	19,525	20	37	6	22	—	13	64	72
Government debt	14,689	79	45	10	13	—	17	—	13
Government deficit	12,814	29	57	10	—	—	—	44	100
Unemployment rate	7,645	60	61	26	38	—	25	—	67
Inflation rate	19,671	48	44	3	45	—	13	41	100
Consumer price index	18,657	51	47	3	48	—	14	45	100
House price index	3,517	—	—	—	—	—	58	—	32
Population	51,930	29	17	4	27	32	5	—	53

Note: This table shows the number of country-year observations in the Global Macro Database (GMD) and the fraction covered in major providers of macroeconomic data (in percent). Our dataset begins in the year 1086. * IFS includes data from both the [IMF MFS](#) and [IMF GFS](#) datasets. † Unemployment data from the World Bank is added from International Labour Organization (ILO). ‡ The GFD coverage is based on the subset of data currently available to us and may not represent the entirety of the GFD dataset.

Figure TA4: Comparing Dataset Coverage by Variable



Note: This figure compares, for each variable, the coverage of the Global Macro Database (GMD) with that of the next most comprehensive source. Blue bars show the number of country-year observations in the Global Macro Database, and the red bars shows the number of country-year observations for the next best source.

Table TA1: Variable Overview

Variable	Abbreviation	Unit	From	To	Forecasts	Countries
A. National accounts						
Nominal GDP	nGDP	Millions of LC	1086	2029	5	229
Real GDP	rGDP	Millions of LC	1270	2029	5	194
Real GDP in USD	rGDP_USD	Millions of USD	1791	2024	—	192
Real GDP per capita	rGDP_pc	LCU per capita	1277	2029	5	194
GDP deflator	deflator	Ratio	1270	2029	5	194
Population	pop	Millions	1277	2030	6	239
B. Consumption and investments						
Real final consumption	rcons	Millions of LC	1800	2024	—	213
Final consumption	cons	Millions of LC	1800	2025	1	219
Final consumption in percent of GDP	cons_GDP	%	1800	2025	1	219
Gross capital formation	inv	Millions of LC	1830	2029	5	218
Gross capital formation in percent of GDP	inv_GDP	%	1830	2029	5	218
Gross fixed capital formation	finv	Millions of LC	1800	2025	1	216
Gross fixed capital formation in percent of GDP	finv_GDP	%	1800	2025	1	216
C. External sectors						
Current account	CA	Millions of LC	1772	2029	5	209
Current account in percent of GDP	CA_GDP	%	1772	2029	5	209
Exports	exports	Millions of LC	1280	2029	5	225
Exports in percent of GDP	exports_GDP	%	1280	2029	5	222
Imports	imports	Millions of LC	1560	2029	5	225
Imports in percent of GDP	imports_GDP	%	1560	2029	5	220
Real effective exchange rate	REER	Index, 2010 = 100	1870	2025	1	180
USD exchange rate	USDfx	1 USD in LC	1791	2025	1	233
D. Government finances						
Government debt	govdebt	Millions of LC	1670	2029	5	197
Government debt in percent of GDP	govdebt_GDP	%	1670	2029	5	197
Government deficit	govdef	Millions of LC	1792	2029	5	200
Government deficit in percent of GDP	govdef_GDP	%	1792	2029	5	200
Government expenditure	govexp	Millions of LC	1722	2029	5	203
Government expenditure in percent of GDP	govexp_GDP	%	1650	2029	5	199
Government revenue	govrev	Millions of LC	1722	2029	5	202
Government revenue in percent of GDP	govrev_GDP	%	1650	2029	5	198
Government tax revenue	govtax	Millions of LC	1750	2024	—	197
Government tax revenue in percent of GDP	govtax_GDP	%	1789	2023	—	190
E. Money and interests						
M0	M0	Millions of LC	1619	2024	—	187
M1	M1	Millions of LC	1841	2024	—	185
M2	M2	Millions of LC	1841	2024	—	183
M3	M3	Millions of LC	1819	2024	—	70
M4	M4	Millions of LC	1870	2020	—	4
Central bank policy rate	cbrate	%	1694	2025	1	167
Short-term interest rate	strate	%	1695	2025	1	140
Long-term interest rate	ltrate	%	1310	2024	—	84
F. Prices and labor market						
Consumer price index	CPI	Index, 2010 = 100	1209	2029	5	214
House price index	HPI	Index, 2010 = 100	1819	2024	—	59
Inflation	infl	%	1210	2029	5	215
Unemployment rate	unemp	%	1760	2029	5	220
G. Financial crises						
Banking crisis dummy	BankingCrisis	Dummy	1800	2020	—	163
Sovereign debt crisis dummy	SovDebtCrisis	Dummy	1800	2020	—	160
Currency crisis dummy	CurrencyCrisis	Dummy	1800	2019	—	160

Note: This table presents the variables included in the Global Macro Database (GMD), along with their abbreviations/variable names, units of measurement, temporal coverage, forecast horizons, and country coverage. “LC” refers to local currency units and “USD” to US dollars.

Table TA2: Comparing the Coverage of Key Macroeconomic Variables

Source	Start Year		Latest		Countries		Variables
	First	Median	Actual	Forecast	Number	Year-Obs.	
GMD	1086	1800	2024	2030	243	55,431	46
GFD	1000	1820	2024	—	236	39,345	37
IFS	1920	1950	2024	—	220	14,179	24
WEO	1960	1970	2024	2029	208	11,086	23
JST	1870	1870	2020	—	18	2,666	29
MAD	1253	1876	2022	—	169	20,590	3
OECD EO	1960	1960	2024	2025	49	2,726	35
PWT	1950	1960	2019	—	183	10,399	3
UN	1970	1970	2020	—	215	10,220	14
WDI	1960	1960	2023	—	222	13,454	36

Note: This table compares the coverage of the key macroeconomic variables included in our dataset across a selected sample of widely-used data providers. See Table [TA1](#) for the list of variables, which also includes derived measures such as GDP-scaled variables. The number of variables refers to the subset of macroeconomic indicators we cover that are available in a given source, not the total number of variables available. The acronyms in the source column refer to the following datasets: GFD = Global Financial Data, IFS = IMF International Financial Statistics, WEO = IMF World Economic Outlook, JST = Jordà-Schularick-Taylor Macro History Database, MAD = Maddison Historical Statistics, OECD EO = OECD Economic Outlook, PWT = Penn World Table, UN = United Nations Statistics, WDI = World Bank World Development Indicators. “Year-Obs.” under the “Countries” header refers to the total number of country-year observations with non-missing information on any of the variables. Note that, for the Maddison dataset, we keep only continuous observations which started in 1253.

TA.3 Sources and Data Access

TA.3.1 Types of data sources

Our dataset combines information from 110 distinct sources that can be categorized along two primary dimensions. The first dimension is update frequency, which distinguishes between current sources (32) that are continuously updated with regular release calendars, primarily from major international organizations and statistical agencies, and historical sources (78) that are either never or infrequently updated without a regular schedule. The second dimension is coverage, where we differentiate between 72 aggregators providing data for multiple countries (e.g., IMF, World Bank) and 38 country-specific sources focused on individual countries, often from national statistical offices or academic research.

Table TA4: Types of Sources in the Global Macro Database

	Current	Historical	Total
Aggregators <i>Examples</i>	24 <i>WEO</i>	48 <i>JST</i>	72
Country-specific <i>Examples</i>	8 <i>FRED</i>	30 <i>Thomas et al. (2010)</i>	38
Total	32	78	110

Note: This table plots information on the number of sources used in the Global Macro Database. We differentiate sources along two dimensions: (1) whether they are continuously updated (*current*) or contain only historical data (*historical*), and (2) whether they report information on several countries (*aggregators*) or only a single country (*country-specific*). Note that we count as historical sources that have been updated on an ad-hoc basis but do not have a clear release calendar, such as the [Jordà et al. \(2017\)](#) Macro History Database.

Table [TA5](#) provides a comprehensive overview of all currently incorporated sources in the Global Macro Database (GMD), along with detailed information about each source. This includes the source reference, the abbreviation used in the GMD, the latest update date, an indication of whether the source is newly digitized, temporal coverage, variable coverage, country coverage, and whether it is classified as a historical source (i.e., a dataset that is not updated or is only infrequently updated).

Table TA5: Dataset Overview

Source	Abbreviation	Updated	Digitized	From	To	Forecasts	Variables	Countries	Historical
Panel A: Aggregator Sources									
Asian Development Bank (2024)	ADB	2024/10/14	No	2000	2023	0	26	49	No
African Development Bank (2018)	AFDB	2025/01/18	No	1980	2020	0	14	53	No
African Union (2024)	AFRISTAT	2024/07/20	No	1990	2023	0	8	22	No
Institute of Economic Research, Hitotsubashi University (2020)	AHSTAT	2024/05/04	No	1860	2013	0	20	6	Yes
European Commission (2024a)	AMECO	2025/01/18	No	1960	2025	1	16	43	No
Arab Monetary Fund (2024)	AMF	2024/07/10	No	1971	2021	0	18	22	No
Barro and Ursúa (2012)	BARRO	2024/05/04	No	1800	2009	0	2	42	Yes
Banque Centrale des Etats de l’Afrique de l’Ouest (2024)	BCEAO	2025/01/18	No	1960	2024	0	23	8	No
Broadberry and Gardner (2022)	BG	2024/09/30	No	1885	2008	0	2	8	Yes
Bank for International Settlements (2024c)	BIS CPI	2025/01/18	No	1661	2023	0	1	62	No
Bank for International Settlements (2024f)	BIS HPI	2025/01/18	No	1927	2024	0	1	58	No
Bank for International Settlements (2024d)	BIS REER	2025/01/18	No	1994	2024	0	1	63	No
Bank for International Settlements (2024a)	BIS USDfx	2025/01/18	No	1791	2024	0	1	189	No
Bank for International Settlements (2024b)	BIS cbrate	2025/01/18	No	1945	2024	0	1	45	No
Bank for International Settlements (2024e)	BIS infl	2025/01/18	No	1662	2023	0	1	62	No
Banca d’Italia (2024)	BIT	2025/01/10	No	1955	2024	0	1	9	Yes
Bordo et al. (2001)	BORDO	2024/04/21	No	1880	1997	0	9	56	Yes
Darvas (2021)	BRUEGEL	2024/04/21	No	1960	2023	0	1	178	Yes
Baron et al. (2020b)	BVX	2024/04/21	No	1870	2016	0	7	48	Yes
United Nations (2024a)	CEPAC	2024/07/11	No	1950	2023	0	23	36	No
International Institute of Social History (2024)	CLIO	2025/01/10	No	1727	2011	0	1	42	Yes
Mack and Martínez-García (2011)	DALLASFED	2024/04/21	No	1975	2023	0	2	26	Yes
University of California – Davis (2024b)	Davis	2024/06/01	No	1818	2012	0	1	55	Yes

Continued on next page

Continued from previous page

Source	Abbreviation	Update	Digitized	From	To	Forecasts	Variables	Countries	Historical
European Commission (2024b)	EUS	2025/01/18	No	1960	2024	0	24	50	No
Food and Agriculture Organization of the United Nations (2024)	FAO	2024/12/10	No	1970	2023	0	3	214	Yes
University of California – Davis (2024a)	FLORA	2024/06/01	No	1799	1975	0	2	12	Yes
Banque de France (2024b)	FRANC ZONE	2025/01/18	No	1991	2019	0	6	15	Yes
Flandreau and Zumer (2009)	FZ	2024/06/24	No	1880	1913	0	14	16	Yes
Smits et al. (2009)	GNA	2024/06/28	No	1800	2005	0	3	17	Yes
Gapminder (2024)	Gapminder	2024/08/14	No	1800	2030	6	1	197	Yes
Grimm (2024)	Grimm	2024/10/30	No	1945	2023	0	1	166	Yes
Schuler (2015)	HFS	2024/06/28	No	1800	2008	0	30	64	Yes
Homer and Sylla (1996)	Homer Sylla	2024/10/28	No	1798	1989	0	3	26	Yes
Ellison et al. (2024)	IHD	2024/10/07	No	1925	1936	0	7	39	Yes
International Labour Organization (2024)	ILO	2024/12/10	No	2000	2023	0	1	215	Yes
Mauro et al. (2015)	IMF FPP	2024/10/28	No	1800	2022	0	4	151	Yes
Mbaye et al. (2018)	IMF GDD	2024/04/21	No	1950	2018	0	3	185	Yes
International Monetary Fund (2024a)	IMF GFS	2025/01/18	No	1972	2023	0	16	158	No
International Monetary Fund (2010)	IMF HDD	2024/06/16	No	1800	2015	0	1	188	Yes
International Monetary Fund (2024b)	IFS	2025/01/18	No	1920	2024	0	19	222	No
International Monetary Fund (2024c)	IMF MFS	2025/01/18	No	1950	2023	0	7	162	No
International Monetary Fund (2024d)	WEO	2025/01/18	No	1960	2029	5	20	208	No
Albers et al. (2023)	JERVEN	2024/10/07	No	1890	2015	0	4	50	Yes
Jones and Obstfeld. (1997)	JO	2024/06/07	No	1850	1945	0	5	13	Yes
Jordà et al. (2017)	JST	2024/06/25	No	1870	2020	0	26	18	Yes
Ljungberg (2019)	LUND	2024/10/14	No	1870	2016	0	1	27	Yes
Laeven and Valencia (2013)	LV	2024/10/17	No	1970	2017	0	4	155	Yes
Inklaar et al. (2018)	MAD	2024/04/04	No	1253	2022	0	3	169	Yes

Continued on next page

Continued from previous page

Source	Abbreviation	Update	Digitized	From	To	Forecasts	Variables	Countries	Historical
Cox and Dincecco (2021)	MD	2024/06/20	No	1650	2010	0	6	31	Yes
Bértola and Rey (2018)	MOXLAD	2024/04/21	No	1870	2010	0	7	20	Yes
Officer and Williamson (2024)	MW	2024/10/02	No	1209	2023	0	10	41	Yes
Mitchell (2013)	Mitchell	2024/04/21	Yes	1750	2010	0	19	142	Yes
National Bank of Serbia (2024)	NBS	2024/06/24	No	1833	1950	0	23	8	Yes
Organisation for Economic Co-operation and Development (2024a)	OECD EO	2025/01/18	No	1960	2025	1	30	49	No
Organisation for Economic Co-operation and Development (2024b)	OECD HPI	2025/01/18	No	1960	2023	0	2	47	No
Organisation for Economic Co-operation and Development (2024c)	OECD KEI	2025/01/18	No	1914	2023	0	6	47	No
Organisation for Economic Co-operation and Development (2024d)	OECD MEI	2025/01/18	No	1935	2023	0	5	47	No
Organisation for Economic Co-operation and Development (1986)	OECD MEI ARC	2024/10/30	No	1955	1984	0	3	18	No
Organisation for Economic Co-operation and Development (2024e)	OECD QNA	2025/01/18	No	1947	2023	0	5	48	No
Organisation for Economic Co-operation and Development (2024f)	OECD REV	2025/01/18	No	1970	2022	0	1	38	No
Feenstra et al. (2015)	PWT	2024/04/21	No	1950	2019	0	3	183	Yes
Reinhart and Rogoff (2009)	RR	2024/04/21	No	1719	2016	0	5	71	Yes
Reinhart and Rogoff (2010)	RR debt	2024/04/21	No	1719	2010	0	1	68	Yes
Schmelzing (2019)	Schmelzing	2024/07/10	No	1310	2018	0	1	8	Yes
Albers (2018)	TH ID	2024/10/07	No	1925	1936	0	2	28	Yes
Federico and Tena-Junguito (2019)	Tena	2024/09/25	No	1800	1938	0	6	150	Yes
United Nations (2024b)	UN	2025/01/18	No	1950	2020	0	8	239	No
Ha et al. (2023)	WB CC	2024/09/05	No	1970	2023	0	2	207	Yes
World Bank (2024)	WDI	2025/01/18	No	1960	2023	0	30	223	No
World Bank (1999)	WDI ARC	2024/07/24	No	1960	1997	0	33	209	Yes
Panel B: Country Specific Sources									
Nakamura and Zarazaga (2001)	ARG 1	2024/06/24	Yes	1901	1935	0	1	1	Yes

Continued on next page

Continued from previous page

Source	Abbreviation	Update	Digitized	From	To	Forecasts	Variables	Countries	Historical
Ministerio de Economia de la Republica Argentina (2024)	ARG 2	2024/06/24	No	1940	2023	0	6	1	Yes
Hutchinson and Ploeckl (2024)	AUS 1	2024/04/21	No	1789	2020	0	6	1	Yes
Vamplew (1987)	AUS 2	2024/07/04	Yes	1788	1917	0	13	1	Yes
Schulze (2000)	AUT 1	2024/04/21	No	1870	1913	0	2	1	Yes
Instituto de Pesquisa Econômica Aplicada (2024)	BRA 1	2024/10/08	No	1872	2023	0	3	1	Yes
Statistics Canada (2024)	CAN 1	2024/10/02	No	1867	1977	0	19	1	Yes
Swiss National Bank (2009)	CHE 1	2024/06/26	No	1907	2005	0	5	1	Yes
Historical Statistics of Switzerland (2012)	CHE 2	2024/10/02	No	1851	1992	0	7	1	Yes
Abildgren (2017)	DNK 1	2024/06/24	No	1487	2023	0	14	1	Yes
Bank of Algeria (2023)	DZA 1	2024/07/24	No	1974	2023	0	3	1	Yes
Instituto Nacional de Estadística (2024)	ESP 1	2025/01/18	No	1995	2023	0	6	1	No
Banco de España (2024)	ESP 2	2024/12/29	No	1277	2014	0	11	1	Yes
Banque de France (2024a)	FRA 1	2025/01/18	No	1970	2024	0	3	1	No
Levy-Garboua and Monnet (2016)	FRA 2	2024/09/30	No	1800	2015	0	3	1	Yes
Thomas et al. (2010)	GBR 1	2024/06/18	No	1086	2016	0	17	1	Yes
Bank Indonesia (2023)	IDN 1	2025/01/18	No	2008	2018	0	14	1	No
Statistics Iceland (1997b)	ISL 1	2024/04/21	Yes	1870	2016	0	3	1	Yes
Statistics Iceland (1997a)	ISL 2	2024/09/26	No	1625	1990	0	28	1	Yes
Baffigi (2013)	ITA 1	2024/07/04	No	1861	2011	0	9	1	Yes
Piselli and Vercelli (2023)	ITA 2	2024/07/04	No	1861	2016	0	14	1	Yes
Istituto Nazionale di Statistica (2024)	ITA 3	2025/01/18	No	1995	2023	0	7	1	No
Bank of Japan (2024)	JPN 1	2024/10/08	No	1882	2017	0	7	1	Yes
Cha et al. (2022)	KOR 1	2024/05/13	No	1911	2016	0	4	1	Yes
Gardner (2022)	LBR 1	2024/09/30	No	1845	1979	0	7	1	Yes
Bank Al-Maghrib (2001)	MAR 1	2024/07/24	No	1985	2024	0	4	1	Yes

Continued on next page

Continued from previous page

Source	Abbreviation	Update	Digitized	From	To	Forecasts	Variables	Countries	Historical
Grytten (2022)	NOR 1	2024/09/25	No	1816	2019	0	3	1	Yes
Eitrheim et al. (2023)	NOR 2	2024/07/08	No	1516	2022	0	15	1	Yes
Statistics Poland (2024)	POL 1	2025/01/18	No	1989	2020	0	15	1	No
Instituto Nacional de Estadística (2001)	PRT 1	2024/07/08	Yes	1549	1998	0	12	1	Yes
Saudi Central Bank (2024)	SAU 1	2025/01/18	No	1973	2020	0	7	1	No
Schön and Krantz (2017)	SWE 1	2024/04/21	No	1290	2020	0	20	1	Yes
Central Bank of the Republic of Türkiye (2024)	TUR 1	2025/01/18	No	1994	2024	0	1	1	No
National Statistics, Republic of China (Taiwan) (2024)	TWN 1	2024/05/13	No	1951	2021	0	4	1	Yes
Federal Reserve Bank of St. Louis (2024)	FRED	2025/01/18	No	1929	2024	0	22	1	No
Carter et al. (2006)	USA 2	2024/09/26	No	1774	2003	0	19	1	Yes
South African Reserve Bank (2024)	ZAF 1	2025/01/18	No	1959	2020	0	17	1	No

TA.3.2 How we update the data

Automated downloads

For current sources, we employ automated data collection through APIs and structured web queries. Our data downloading and processing pipeline has the following key functions:

- Automatic downloading and processing new data as it becomes available
- Data validation and harmonization
- Integration of new information into the existing dataset

Manual collection

Historical sources require careful manual processing. This involves the one-time collection of historical datasets, digitization of printed materials where necessary, standardization of formats and units, and integration with the automated updating system for current data. This meticulous process ensures that the historical data maintains the same quality standards as our automated collections.

TA.3.3 Data storage and version control

Raw data storage

A critical feature for combining many dozens of historical sources with up-to-date recent data is to maintain a comprehensive archive of previous versions. For every source file, we always maintain the original data files in their native formats (csv, Excel, Stata, etc.) at the time they were downloaded. We also maintain any additional source documentation or meta data, as available. If the data is newly digitized or comes from an unusual source, we store the PDF files and archive the relevant web pages. In all cases, we always record the time stamp when a dataset was downloaded and the source URLs.

Version control

Our version control system tracks data vintages from each source. We store newly-downloaded incremental updates and always record whether data values are genuinely new information or represent revisions (i.e., updates of data points that were already previously available). As such, we maintain complete historical records of all data points, and document

all modifications and harmonization steps. This systematic approach ensures full traceability of every transformation made to the original data.

TA.3.4 Data update process

Update monitoring and integration

Our update process begins with a continuous monitoring of current sources for new releases, supported by an automated notification system for new data availability and regular checks of historical sources for potential updates. The integration phase involves automated comparison with existing data, identification of revised values, integration of new data points, and maintenance of vintage data.

Quality control

Each update undergoes a rigorous quality control process consistent with the initial release of the Global Macro Database. Among others, we always run (a) a set of automated data validity checks, (b) unit and format verification checks, (c) consistency checks relative to previous versions, and (d) manually review all significant changes. This process is designed to make the resulting dataset as error-free as possible.

In addition to these initial checks, we always manually check the time series plot for each country and variable for potential issues.

TA.4 Variable Definitions

TA.4.1 Statistical framework

Wherever possible, our variable definitions follow the System of National Accounts 2008 (SNA 2008), the international statistical standard for national accounts adopted by the United Nations Statistical Commission. The SNA 2008 provides a comprehensive, consistent, and flexible framework for collecting and reporting macroeconomic statistics.

The SNA framework is reasonable starting point for our dataset because it provides internationally agreed-upon concepts, definitions, and classifications, ensures consistency and comparability across countries and time periods as much as possible, and offers standardized methodologies for data compilation.

For variables that are not reported as part of national accounts statistics, we try to adopt a consensus approach by surveying the meta data and best practices in existing work.

TA.4.2 Main national account variables

Nominal gross domestic product (nGDP)

Nominal GDP is defined as the sum of gross value added by all resident producer units plus the sum of taxes on products less subsidies on products ([United Nations, 2009, Chapter 16, Section C](#)).

nGDP is measured at current market prices in millions of local currency units.

Real gross domestic product (rGDP)

Real GDP is a volume measure of GDP that removes the effects of price changes over time ([United Nations, 2009, Chapter 15, Section C](#)).

rGDP is based on the prices of a reference year and measured in millions of local currency units at constant prices.

rGDP_USD is based on the prices of a reference year and measured in millions of US dollars at constant prices.

rGDP_pc is based on the prices of a reference year and measured in millions of local currency units per capita at constant prices.

GDP deflator (deflator)

The GDP deflator is defined as a price index derived by dividing nominal GDP by real GDP ([United Nations, 2009, Chapter 15, Section C](#)).

deflator measures the overall price level of all domestically produced final goods and services. The deflator is set to 100 in the base year and expressed as an index.

Population (pop)

Population is the total number of persons present in the economic territory, including both nationals and foreigners ([United Nations, 2009, Chapter 19, Section B](#)).

pop is measured in millions of persons.

TA.4.3 Consumption and investment

Final consumption (cons)

Final consumption is the total expenditure on the consumption of goods and services ([United Nations, 2009, Chapter 9, Section D](#)).

cons is measured in millions of local currency units.

cons_GDP is based on the prices of a reference year and measured and measured in millions of local currency units in percentage of nominal GDP.

TA.4.4 Investment

Gross capital formation (inv)

Gross capital formation is measured by the total value of the gross fixed capital formation, changes in inventories and acquisitions less disposals of valuables. ([United Nations, 2009, Chapter 10, Section B](#)).

inv is measured in millions of local currency units.

inv_GDP is measured in millions of local currency units in percentage of nominal GDP.

Gross fixed capital formation (finv)

Gross fixed capital formation is measured by the total value of a producer's acquisitions, less disposals, of fixed assets during the accounting period plus certain specified expenditure on services that adds to the value of non-produced assets. ([United Nations, 2009, Chapter 10, Section B](#)).

finv is measured in millions of local currency units.

finv_GDP is measured in millions of local currency units in percentage of nominal GDP.

TA.4.5 External sector

Exports (exports)

Exports are goods and services produced in one economy and sold to another economy, valued free on board (f.o.b.) at the border of the exporting country ([United Nations, 2009, Chapter 15, Section B](#)).

exports is measured in millions of local currency units.

exports_GDP is measured in millions of local currency units in percentage of nominal GDP.

Imports (imports)

Imports are goods and services purchased by residents from nonresidents, originally valued cost, insurance, and freight (c.i.f.) but converted to f.o.b. ([United Nations, 2009, Chapter 15, Section B](#)).

imports is measured in millions of local currency units.

imports_GDP is measured in millions of local currency units in percentage of nominal GDP.

Current account balance (CA)

The current account is defined as the sum of the balances on goods, services, primary income, and secondary income ([United Nations, 2009, Chapter 16, Section B](#)). It is considered a key indicator of an economy's saving-investment relationship with the rest of the world.

CA is measured in millions of local currency units.

CA_GDP is measured in millions of local currency units in percentage of nominal GDP.

Real effective exchange rate (REER)

The real effective exchange rate, or REER, is the trade-weighted average of bilateral exchange rates adjusted for relative price levels.

REER is expressed as an index equal to 100 in the base year and thus measures (changes in) international competitiveness.

USD exchange rate (USD_{fx})

The US dollar exchange rate is defined as the value of one U.S. dollar in terms of local currency units. Wherever possible we report end-of-period rates.

TA.4.6 Government finance

Government debts (govdebt)

Government debt is the total liabilities of the government requiring future payments of interest and/or principal. It includes loans, debt securities, and other borrowings ([United Nations, 2009, Chapter 22, Section D](#)).

govdebt is measured in millions of local currency units.

govdebt_GDP is measured in millions of local currency units in percentage of nominal GDP.

Government revenues (govrev)

Government revenue is the increase in net worth of the government resulting from incoming transactions; it includes taxes, social contributions, grants, and other revenue ([United Nations, 2009, Chapter 22, Section C](#)).

govrev is measured in millions of local currency units.

govrev_GDP is measured in millions of local currency units in percentage of nominal GDP.

Government tax revenue (govtax)

Government tax revenues are compulsory transfers to government units, including taxes on income, profits, goods and services, and international trade ([United Nations, 2009, Chapter 22, Section C](#)).

govtax is measured in millions of local currency units.

govtax_GDP is measured in millions of local currency units in percentage of nominal GDP.

Government expenditure (govexp)

Government expenditure is a decrease in net worth of the government resulting from transactions, including from compensation of employees, use of goods and services, and transfers ([United Nations, 2009, Chapter 22, Section C](#)).

govexp is measured in millions of local currency units.

govexp_GDP is measured in millions of local currency units in percentage of nominal GDP.

Government deficit (govdef)

Government deficit refers to net lending/borrowing by the government, i.e. the difference between revenue and expenditure ([United Nations, 2009, Chapter 22, Section C](#)).

govdef is the primary measure of a country's fiscal position and measured in millions of local currency units.

govdef_GDP is measured in millions of local currency units in percentage of nominal GDP.

TA.4.7 Money and interest rates

Money supply (M0)

M0 is defined as notes and coins in circulation outside depository corporations ([International Monetary Fund, 2016, Chapter 6, Section IV](#)). It is considered the most liquid monetary aggregate.

M0 measured in millions of local currency units.

Money supply (M1)

M1 is defined as currency in circulation plus transferable (demand) deposits ([International Monetary Fund, 2016, Chapter 6, Section III](#)). It is the most liquid monetary aggregate including bank deposits.

M1 is measured in millions of local currency units.

Money supply (M2)

M2 is defined as M1 plus time and savings deposits and includes less liquid monetary assets ([International Monetary Fund, 2016, Chapter 6, Section III](#)).

M2 is measured in millions of local currency units.

Money supply (M3)

M3 is defined as M2 plus marketable instruments issued by depository corporations (refer to [International Monetary Fund, 2016, Chapter 6, Section III](#)). It represents the broadest monetary aggregate measured within the banking system.

M3 is expressed in millions of local currency units.

Money supply (M4)

M4 is defined as M3 plus debt securities issued by the central government that are held by money holders (refer to [International Monetary Fund, 2016, Chapter 6, Section III](#)). It is the most comprehensive measure of the money supply.

M4 is also expressed in millions of local currency units.

Central bank policy rate (cbrate)

The central bank policy rate is the key interest rate used to implement or signal monetary policy stance ([International Monetary Fund, 2016, Chapter 4, Section II](#)). It usually applies

to short-term liquidity operations.

cbrate is measured in percent per annum.

Short-term interest rate (strate)

The short-term interest rate is the market rate on short-term government securities or interbank rates, usually referring to three-months maturity ([International Monetary Fund, 2016, Chapter 4, Section II](#)).

strate is measured in percent per annum.

Long-term interest rate (ltrate)

The long-term interest rate is the market yield on long-term government bonds, usually referring to ten-year maturity ([International Monetary Fund, 2016, Chapter 4, Section II](#)).

ltrate is measured in percent per annum.

TA.4.8 Prices and labor market

Consumer price index (CPI)

The consumer price index (CPI) is a measure of the average change in prices paid by consumers, with the weights based on household consumption patterns ([United Nations, 2009, Chapter 15, Section A](#)).

CPI is expressed as an index equal to 100 in the base year.

House price index (HPI)

The house price index (HPI) measures changes in residential property prices, where possible adjusted for quality ([Eurostat, 2013](#)).

HPI is expressed as an index equal to 100 in the base year.

Inflation (infl)

Inflation is defined as the period-on-period percentage change in the consumer price index, which measures the rate of price level changes ([United Nations, 2009, Chapter 15, Section A](#)).

infl is expressed in percent per annum.

Unemployment rate (unemp)

The unemployment rate is defined as the number of unemployed persons as a percentage of the labor force ([United Nations, 2009, Chapter 19, Section D](#)). Wherever possible, it is based on International Labor Organization standards.

unemp is expressed in percent.

TA.5 Priority Ordering of Data Sources

Our extensive experience working with macroeconomic time series has led us to develop a hierarchy for how different sources should be prioritized. This ordering reflects both the reliability of the data and the practical considerations of maintaining a comprehensive database. The hierarchy follows three main tiers:

TA.5.1 First tier: Modern official sources

Modern official sources, particularly those from international organizations, national statistical offices, and central banks receive the highest priority. These institutions typically have the most accurate and up-to-date information for their respective countries, along with detailed documentation of methodologies and regular revision schedules.

TA.5.2 Second tier: Country-specific historical sources

The second tier consists of country-specific historical sources, often compiled by economic historians or research institutions focusing on particular countries or regions. These sources frequently offer invaluable historical data that has been carefully reconstructed and have often been evaluated through the academic peer review process. Notable examples include the historical statistics for Australia compiled by [Vamplew \(1987\)](#), Portuguese historical statistics by [Instituto Nacional de Estatística \(2001\)](#), and Argentinian long-run series by [Nakamura and Zarazaga \(2001\)](#).

TA.5.3 Third tier: Other aggregators

International aggregators such as the IMF, World Bank, or OECD form the third tier. While these sources provide extensive cross-country coverage and standardized definitions, they sometimes sacrifice historical depth or country-specific accuracy.

This ordering has proven robust through extensive testing and practical application. It balances the trade-offs between data accuracy, historical coverage, methodological consistency, and maintenance feasibility. The hierarchy is not rigid, however, and we document any deviations from this general ordering in our detailed source notes for each country and variable.

TA.6 Measurement Issues

TA.6.1 Data quality

A key contribution of our dataset is the systematic approach to data quality control through comprehensive visual inspection of all time series. For each of our 46 variables and 243 countries/territories, we plot the data from every available source on a single graph, allowing for detailed comparison and anomaly detection. This results in over 4,000 individual plots that we manually inspect for data quality issues.

Visual inspection process

Our visual inspection process creates plots for each country-variable combination. These plots comprehensively display our final GMD estimates, a GMD forecast (where available), data points from all available sources, clear indication of splice points between different sources, and notes on major adjustments or concerns.

This visualization process allows us to identify several critical types of data quality issues. We detect level shifts, which appear as unexpected jumps in the series that might indicate currency changes, definition changes, or data errors. We also identify source discrepancies, where different sources report substantially different values for the same period. The process also reveals outliers that deviate significantly from the series trend, inconsistencies in the units or currency in which a data series is recorded, and splicing problems where different data sources are not correctly “stitched together.”

Documentation and correction

After identifying a potential data quality issue, we implement a systematic correction procedure. We begin by thoroughly documenting the nature of the anomaly and cross-referencing it with other sources and historical events. This investigation helps us determine whether the issue reflects a genuine economic event, a definitional change, or a data error requiring correction. Based on this assessment, we apply the necessary corrections

while maintaining a detailed documentation of all adjustments. Throughout this process, we carefully flag any remaining uncertainties in the dataset to ensure full transparency.

Quality control documentation

We maintain a comprehensive set of PDF documents containing plots for each country and variable, accessible at www.globalmacrodata.com/documentation. This documentation provides extensive details, including a complete source information, the time spans each data source is used for constructing our time series, notes on specific adjustments or concerns, and detailed explanations of any splicing procedures applied.

As an illustrative example, consider the case of government expenditure data in Argentina. We use the following sources:

- [Mitchell \(2013\)](#): 1864-1989
- [International Monetary Fund \(2024a\)](#): 1990-1992
- [International Monetary Fund \(2024d\)](#): 1993-2029

These plots serve dual purposes: they function as both a quality control tool and transparent documentation of our data construction process. This comprehensive approach allows users to evaluate the reliability of specific series and understand precisely how different sources were combined to create our final estimates.

TA.6.2 Dealing with data revisions

A key challenge in maintaining a long-run macroeconomic database is the treatment of data revisions. Statistical offices and international organizations frequently revise their estimates as new information becomes available or methodologies improve. Without proper treatment, these revisions would propagate through the entire historical series when using standard chain-linking procedures, causing implausible changes to historical values.

Fixed year approach

To address the issue of data revisions, we implement a reference year approach that effectively separates historical data from contemporary revisions. We establish 2019 as our reference year for the current version of the dataset, with historical data (pre-2019) chain-linked backwards and contemporary data (post-2019) chain-linked forwards from this base year.

This approach creates a clear framework for handling new data and revisions:

- Values before 2019 remain unchanged
- Values after 2019 incorporate all new information and revisions
- The reference year serves as an anchor point between historical and contemporary data

Going forward, we will occasionally consider a change in the reference year to take into account possible data revisions to 2019 values of the data.

Example

To illustrate our approach of fixing the reference year, consider a hypothetical case for real GDP. We begin with a historical source providing data for 1850-1989 and a contemporary source (e.g., [International Monetary Fund \(2024d\)](#)) covering 1990-2024. When a new ([International Monetary Fund, 2024d](#)) release revises the values for the 2022-2024 period and adds 2025-2029 forecasts, the treatment of these revisions differs significantly under different approaches.

Under traditional chain-linking starting from the most recent data, revisions to recent years would affect the entire series back to 1850, meaning historical values would change despite no new historical information becoming available. In contrast, our reference year approach provides a more stable framework:

- 2019 serves as the fixed reference point
- Data for 1850-2019 is constructed by chain-linking backwards from 2019
- Data for 2020-2025 is constructed by chain-linking forwards from 2019
- Revisions to 2022-2024 and the addition of 2025 only affect post-2019 values

Implementation

For each variable, we implement a systematic procedure that begins with fixing the level of the series in 2019 using our preferred contemporary source. For earlier years ($t < 2019$), we use growth rates from historical sources to chain-link backwards from 2019, with these values remaining fixed unless errors are discovered. For later years ($t > 2019$), we employ growth rates from contemporary sources to chain-link forwards from 2019, allowing these values to update with each new data release.

This methodology ensures that our historical estimates remain stable while still incorporating all relevant contemporary revisions and updates. We plan to periodically update

the reference year (e.g., to most recent years) in the next major releases of the dataset. At this point, the entire series will be recomputed to incorporate any important historical revisions.

TA.6.3 Base years for index variables

Several variables in our dataset are expressed either in constant prices or as indices (e.g., Consumer Price Index, GDP deflator, real GDP). These variables require a common reference year for meaningful comparison. We implement a systematic approach to standardize reference years across all index variables.

Primary reference source

We use the IMF’s World Economic Outlook ([International Monetary Fund, 2024d](#)) as our primary reference source for reference years. This choice is motivated by several key advantages: the WEO offers wide country coverage, maintains a regular updating schedule, employs consistent methodological treatment, and enjoys broad usage in applied research.

Rebasing methodology

Our rebasing procedure follows a hierarchical approach with two main pathways. The first is direct WEO alignment, applied when a series overlaps with WEO data. In these cases, we identify the overlapping period between the source and WEO, calculate a rebasing factor using this period, and apply the factor to the entire source series. The second pathway, an indirect WEO alignment, is employed when a series has no overlap with the WEO data. This requires identifying an intermediate source with WEO overlap, rebasing that intermediate source to the WEO reference year, and finally using the overlapping period with the historical source to apply a chain of rebasing factors.

Mathematical implementation

For a series from source A that needs to be rebased to match source B (e.g., WEO), we compute the rebased values as follows:

$$X_t^{new} = X_t^A \times \frac{\overline{X_s^B}}{\overline{X_s^A}} \quad (\text{TA1})$$

where X_t^{new} is the rebased value at time t , X_t^A is the original value from source A at time t , $\overline{X_s^B}$ is the mean of source B values in the overlapping period s , and $\overline{X_s^A}$ is the mean of source A values in the overlapping period s .

This simple ratio approach ensures three key properties: it preserves the relative changes in the original series, adjusts the levels to match the reference source in the overlapping period, and can be easily chained when intermediate sources are needed. For example, if a historical source overlaps with WEO data during 1990-2000, we calculate the mean of both series for this period, compute their ratio, and multiply the entire historical series by this ratio.

Special cases and quality control

For cases where neither direct nor indirect WEO alignment is possible, we proceed on a case-by-case basis, employing alternative strategies to maintain data quality. We turn to other major international sources, such as the World Bank or OECD, as reference points. Each case requires careful documentation of our rebasing assumptions in country-specific notes, and we explicitly flag these series to indicate higher uncertainty in the rebasing process.

To ensure accurate rebasing across all cases, we implement a comprehensive set of quality control measures. We visually inspect the rebased series, verify that the growth rates were correctly preserved, cross-validate with alternative sources as available, and document of all rebasing factors. This standardization ensures that all index variables in our dataset share a common reference year, facilitating cross-country and cross-time comparisons while maintaining consistency with contemporary international statistics.

TA.6.4 Changes in currency

The long-run nature of our dataset requires careful handling of currency changes and redenominations. Historical sources often report values in currencies that were in use at the time of recording, while modern sources typically adjust historical values to current currencies. We identify and address two distinct types of currency changes:

1. Adoption of new currencies: Complete changes in the monetary unit, such as:

- Introduction of new national currencies post-independence
- Currency union adoptions (e.g., Euro)
- Post-hyperinflation currency reforms

2. **Redenomination of existing currencies:** Technical adjustments to the same basic monetary unit, typically:

- Removal of zeros after hyperinflation periods
- Technical currency reforms maintaining the same basic unit

For new currency adoptions, we apply the official conversion rate at the time of change. For example, in the case of Euro adoption, we use the irrevocable exchange rates set by the European Central Bank (e.g., 1.95583 Deutsche Mark = 1 Euro for Germany). All historical values are converted using these official rates, and we document the conversion rate and date in country-specific notes.

Currency redenominations, particularly following periods of hyperinflation, require special attention. For instance, Turkey removed six zeros from the Lira in 2005, requiring all pre-2005 values to be divided by 1,000,000 to maintain consistency. Brazil presents a more complex case, with multiple changes between 1942 and 1994:

- 1942-1967: Cruzeiro
- 1967-1970: Cruzeiro Novo, 1000:1 redenomination
- 1970-1986: Cruzeiro
- 1986-1989: Cruzado, 1000:1 redenomination
- 1989-1990: Cruzado Novo, 1000:1 redenomination
- 1990-1993: Cruzeiro
- 1993-1994: Cruzeiro Real, 1000:1 redenomination
- 1994-present: Real, 2750:1 conversion

For such complex cases, we implement a systematic procedure in four steps. First, we document the complete chain of currency changes. Second, we apply conversions sequentially using official exchange rates. Third, we cross-validate the results with multiple sources when available. Fourth, we record potential measurement uncertainties in periods of extreme inflation.

To ensure accurate currency conversions, we employ several quality control measures. These start with an automated detection of unusual jumps in series around known currency change dates and a cross-validation with multiple sources where available. Most importantly, we manually review all major currency changes.

For each country-year observation affected by currency changes, we maintain detailed documentation of the original currency and value, applied conversion rate(s), and final adjusted value. This information is available in the country-specific technical notes. In cases where sources disagree about the appropriate conversion rate, particularly during periods of high inflation, we use geometric averages of available estimates and flag these observations for higher uncertainty.

TA.6.5 Dealing with series breaks

Series breaks occur when the same variable exhibits a discontinuity due to methodological changes, redefinitions, or changes in source data. We identify a series break when either:

- There is an explicit documentation of methodological change
- There is an unexplained jump in the series that unambiguously cannot be attributed to economic events
- Other sources than the focal one report substantially different growth rates or values for the same period

We employ two distinct methodologies for addressing series breaks, depending on the availability of overlapping data.

Method 1: Overlapping data available

When we have overlapping observations between two sources at the break point t , we adjust the historical series using the ratio method:

$$X_s^{adjusted} = X_s^{old} \times \frac{X_t^{new}}{X_t^{old}} \quad \text{for all } s < t \quad (\text{TA2})$$

where:

- $X_s^{adjusted}$ is the adjusted value for period s
- X_s^{old} is the original value from the old series
- X_t^{new} is the value from the new series at the break point
- X_t^{old} is the value from the old series at the break point

Method 2: No overlapping data

When there is no overlapping data at the break point, we use growth rates from both series around the break to construct an adjustment factor. Let t be the break year. We:

1. Calculate median growth rates for three years before break ($t - 3, t - 1$):

$$g_1 = \text{median} \left(\frac{X_s^{\text{old}} - X_{s-1}^{\text{old}}}{X_{s-1}^{\text{old}}} \right) \quad \text{for } s \in \{t - 3, t - 2, t - 1\} \quad (\text{TA3})$$

2. Calculate median growth rates for three years after break ($t, t + 2$):

$$g_2 = \text{median} \left(\frac{X_s^{\text{new}} - X_{s-1}^{\text{new}}}{X_{s-1}^{\text{new}}} \right) \quad \text{for } s \in \{t + 1, t + 2, t + 3\} \quad (\text{TA4})$$

3. Compute adjustment factor:

$$\theta = \frac{1 + g_2}{1 + g_1} \quad (\text{TA5})$$

4. Apply adjustment to historical series:

$$X_s^{\text{adjusted}} = X_s^{\text{old}} \times \theta \quad \text{for all } s < t \quad (\text{TA6})$$

The final combined series is then:

$$X_t = \begin{cases} X_t^{\text{adjusted}} & \text{if } t < \text{break year} \\ X_t^{\text{new}} & \text{if } t \geq \text{break year} \end{cases} \quad (\text{TA7})$$

Implementation notes

We strictly give preferences to Method 1 whenever overlapping data is available. Method 2 is only used when no overlapping observations exist. When multiple breaks exist in a series, we apply these methods sequentially. Crucially, all series breaks and adjustment methods are documented in country-specific notes.

TA.6.6 Doubtful data points

In compiling a comprehensive macroeconomic database, we systematically identify and address doubtful data points that may reflect measurement errors rather than genuine

economic phenomena. We classify these into two main categories: outliers and definitionally impossible values.

Outliers

We define outliers as observations that exhibit implausible changes in the level of a series. Our identification process combines economic validation and source comparison. For economic validation, we cross-reference flagged observations with other sources, verify if large changes can be explained by known economic shocks, and compare with related economic variables for consistency. The source comparison involves checking alternative data sources for the same period and evaluating the reliability of different sources.

When outliers are identified, we follow a three-step protocol. First, we replace values with another source if a more reliable source is available. Second, if no reliable alternative exists, we set the observation to missing. Third, all adjustments are meticulously recorded in the country-specific notes.

Definitionally impossible values

Certain variables have natural constraints on their possible values based on economic definitions. We systematically check for and address two types of constraints. First, sign restrictions dictate that stock variables (e.g., money supply, population), price indices, and exchange rates must be positive, while flow variables (e.g., government deficit, current account) can be negative. Second, logical bounds require that percentages (e.g., unemployment rate) must lie between 0 and 100, and ratios (e.g., debt-to-GDP) must be non-negative.

When encountering impossible values, we implement a three-step verification process. First, we compare the values with alternative sources to identify correct values. Second, we check for potential unit or sign errors in the source data. Third, we investigate possible definitional changes or special circumstances that might give rise to unusual values.

Quality control process

For both types of doubtful data points, we implement a systematic review process:

1. Automated flagging of potential issues
2. Manual review of flagged observations
3. Cross-validation with multiple sources

4. Documentation of all adjustments
5. Regular review of flagging thresholds

All identified doubtful data points and their resolution are recorded in our technical documentation, allowing users to assess the reliability of specific observations and understand any adjustments made to the raw data.