The Stata Journal (2026) *vv*, Number *ii*, pp. 1–12

gmd: The Easy Way to Access the World's Most Comprehensive Macroeconomic Database

Mohamed Lehbib
Department of Finance
National University of Singapore
Singapore
lehbib@u.nus.edu

Karsten Müller
Department of Finance and RMI
National University of Singapore and CEPR
Singapore
kmueller@nus.edu.sg

Abstract. This article introduces `gmd`, a command to access the Global Macroeconomic Database (GMD) introduced by Müller et al. (2025), which is a long-run, open-source [1], and continuously updated database designed to unify and extend existing macroeconomic sources. The command provides access to the latest release of the GMD, previous vintages, harmonized data from dozens of other providers such as the World Bank, the International Monetary Fund, and Eurostat, and some other utility functions.

Keywords: st0001, GMD, economic data, historical data, reproducible research

1 Introduction

Researchers interested in cross-country macroeconomic data face a common dilemma. The data they rely on is scattered across many dozens of providers, but there is no single source that integrates and harmonizes even the most widely-used variables. As a result, there are essentially two choices: (a) use data from a single source, which may be incomplete or lack important data points, or (b) spend weeks or months trying to combine data from different sources before starting any analysis.

To illustrate the challenge, consider what an individual researcher would have to go through to create a truly comprehensive, harmonized macroeconomic dataset covering as many countries as possible. The first challenge is to find and retrieve the raw data from dozens or even hundreds of sources. Downloading time series from the World Bank or International Monetary Fund is easy enough, but other data is stored in irregularly formatted Excel files, or only available in physical books, scanned PDFs, or CD-ROMs. Some of these data may have to be digitized.

Once that hurdle is cleared, the data needs to be read in and cleaned. The reporting units of each variable, however, differ widely across datasets. These units have to be harmonized and indices re-based. Often, currencies need to be converted—a non-trivial task when working with very long time periods, because currency reforms are often incompletely incorporated even by leading data providers.

Even after going through this arduous process, it turns out there are in fact many errors in the raw data. These need to be manually identified and rectified. And then the

1. To access the project source-code, please visit: github.com/KMueller-Lab/Global-Macro-Database.

question arises of how one should combine the data from many different sources that may at times disagree, given that simply taking the data as-is will create statistical breaks, and thus trends that never actually appeared in the raw data.

The Global Macro Database (GMD) [2], introduced by Müller et al. (2025), is designed to overcome these challenges by providing the most comprehensive database of macroeconomic statistics—harmonized, ready-to-use, and transparently documented. The key novelty relative to existing work is that the GMD is updated quarterly, which is made possible by the use of an automated pipeline that integrates more than 100 historical and modern sources. At the time of writing, the GMD has been downloaded more than 11,000 times by researchers from all over the globe.

To make access to these comprehensive time series as easy as possible, this article describes the `gmd` [3] command. Using `gmd` allows users not only to download the most recent version of the dataset. It also provides access to previous vintages—an important factor for making research using it fully reproducible—as well as the cleaned raw data from each source. The latter is particularly important: it allows users to easily access up-to-date data from more than 100 sources without having to manually clean it, given that the underlying code is public on the GMD GitHub repo. In fact, many users already use the GMD precisely for comparing data points from different providers.

The rest of this article proceeds as follows. Section 2 provides more details on the Global Macro Database and why it is useful in applied research. Section 3 introduces the `gmd` command and its functionalities. Section 4 illustrates how easy it is to work with macroeconomic data using the `gmd` command using two canonical examples. Section 5 concludes.

2 The Global Macroeconomic Database

The GMD uses 111 sources as inputs. These range from datasets produced by international organizations to country-specific sources that were often compiled by economic historians. Equipped with these sources, the GMD cleans and harmonizes the data, and addresses any errors. Once the source data are cleaned, the GMD splices ready-to-use time series that adjust for statistical breaks.

In the current release (2025_12), 74 variables are included, which span 240 countries and territories over the time period 1086 to 2024 with projections up to 2030. Unlike other data efforts, the GMD is centred around an automated pipeline that continuously integrates the latest data points and revisions from modern sources and combines them to construct the longest possible series.

This section provides a high-level overview of the GMD. For details, including technical specifics, please see Müller et al. (2025) and the technical documentation available at www.globalmacrodata.com.

2. The project website is: globalmacrodata.com.
3. To access the command source-code, please visit: github.com/KMueller-Lab/Global-Macro-Database-Stata.

2.1 Variables

Table 1 contains an overview of the variables included in the GMD, including the variable codes, reporting units, and coverage. The GMD comprises what we would consider the most frequently used macroeconomic indicators.

Table 1: Variable Overview

Variable	Abbreviation	Derived	Range	Forecast	Country
A. National accounts					
Nominal GDP	`nGDP`	USD	1086-2029	4	230
Real GDP	`rGDP`	USD, pc	1270-2029	4	227
Real GDP in USD	`rGDP_USD`	pc	1791-2025	—	218
GDP deflator	`deflator`		1270-2029	4	227
Final consumption	`cons`	USD, GDP	1800-2027	2	220
Gross capital formation	`inv`	USD, GDP	1830-2029	4	219
Gross fixed capital formation	`finv`	USD, GDP	1800-2027	2	217
B. External sectors					
Current account , % GDP	`CA_GDP`	LC, USD	1772-2029	4	207
Exports	`exports`	USD, GDP	1280-2029	4	226
Imports	`imports`	USD, GDP	1560-2029	4	226
Real effective exchange rate	`REER`		1870-2026	1	180
USD exchange rate	`USDfx`		1791-2029	4	236
C. Consolidated government finances					
Gov. debt , % GDP	`govdebt_GDP`	LC	1670-2025	—	185
Gov. deficit , % GDP	`govdef_GDP`	LC	1689-2029	4	198
Gov. expenditure , % GDP	`govexp_GDP`	LC	1689-2024	—	173
Gov. revenue , % GDP	`govrev_GDP`	LC	1689-2024	—	170
Gov. tax revenue , % GDP	`govtax_GDP`	LC	1789-2024	—	147
D. General government finances					
Gov. debt , % GDP	`gen_govdebt_GDP`	LC	1800-2029	4	195
Gov. deficit , % GDP	`gen_govdef_GDP`	LC	1800-2029	4	197
Gov. expenditure , % GDP	`gen_govexp_GDP`	LC	1800-2029	4	198
Gov. revenue , % GDP	`gen_govrev_GDP`	LC	1800-2029	4	197
Gov. tax revenue , % GDP	`gen_govtax_GDP`	LC	1850-2027	2	148
E. Central government finances					
Gov. debt , % GDP	`cgovdebt_GDP`	LC	1670-2025	—	185
Gov. deficit , % GDP	`cgovdef_GDP`	LC	1689-2024	—	167
Gov. expenditure , % GDP	`cgovexp_GDP`	LC	1689-2024	—	173
Gov. revenue , % GDP	`cgovrev_GDP`	LC	1689-2024	—	170
Gov. tax revenue , % GDP	`cgovtax_GDP`	LC	1789-2024	—	147
F. Money and interests					

Continued on next page

Continued from previous page

Variable	Abbreviation	Derived	Range	Forecast	Country
M0	`M0`		1619-2025	—	174
M1	`M1`		1841-2025	—	165
M2	`M2`		1841-2025	—	154
M3	`M3`		1819-2025	—	164
M4	`M4`		1870-2020	—	4
Central bank policy rate	`cbrate`		1694-2026	1	167
Short-term interest rate	`strate`		1695-2026	1	136
Long-term interest rate	`ltrate`		1727-2026	1	83
G. Prices, labor market, and population					
Consumer price index	`CPI`		1209-2029	4	216
House price index	`HPI`		1819-2025	—	58
Inflation	`infl`		1210-2029	4	218
Unemployment rate	`unemp`		1760-2029	4	220
Population	`pop`		1277-2030	5	235
H. Financial crisis dummy					
Banking crisis	`BankingCrisis`		1800-2020	—	163
Sovereign debt crisis	`SovDebtCrisis`		1800-2020	—	160
Currency crisis	`CurrencyCrisis`		1800-2019	—	160

Note: This table presents the variables included in the Global Macro Database (GMD), along with their abbreviations/variable names, available derivations, temporal coverage, forecast horizons, and country coverage. Except for variables explicitly in USD, all variables in levels are recorded in millions of local currency units. Indices (CPI, HPI) are recorded such that 2015 = 100. Interest rates, inflation, and unemployment are recorded in percent (e.g., 5 means 5%). Derived refer to other versions of the variables available in the data, where "LC" stands for local currency (instead of % of GDP), "USD" for US dollars, and "GDP" for fraction of GDP.

2.2 Data collection

The GMD is based on two broad categories of sources:

1. *Modern sources*: Datasets from major international organizations (e.g., Eurostat, OECD) that are continuously updated. The GMD retrieves these via an automated code pipeline and APIs to ensure continuous updates.
2. *Historical sources*: Academic work or datasets from national authorities (e.g., Mitchell (2013) or Inklaar et al. (2018)) that are updated infrequently or cannot be accessed through our automated pipeline. Integrating these sources often requires manual digitization and extensive harmonization of data from archival records.

This classification showcases the GMD's core advantages: it covers both the historical versions that need a one-time, albeit tedious, processing, and the modern sources

which provide updates and continuously extend the database.

2.3 Data cleaning

As outlined in the introduction, macroeconomic data from different sources cannot be easily combined with each other, for a variety of reasons.

Currency issues: A harmonized time series requires that all historical observations to be expressed in the currency units a country currently uses. This ensures that observations are consistent over time and not distorted by currency reforms, with the introduction of the Euro being a prominent example.

Units issues: Inconsistent reporting of units is a major challenge. Some sources switch units without necessarily providing the adequate documentation. In the GMD, we ensure that each variable is expressed in the units we assign to it. All levels variables are expressed in millions of local currency, all rates in percent, and indices are set equal to 100 in a common base year.

Base year issues: For variables expressed in constant prices (e.g., real GDP), sources use different base years, making a cross-country comparison impossible. To resolve this, we identify the implicit base year in the raw source and construct variables expressed in 2015 prices.

Statistical breaks: A statistical break refers to a jump in the time series resulting from changes in methodology, definitions, or coverage rather than economic fundamentals. We systematically identify these discontinuities and correct them using a ratio-splicing approach, which aligns the levels of the historical series to the modern series while preserving the historical growth rates.

2.4 Data construction

Once we ensure that the data reported for each variable follow consistent standards, currencies, and units, we combine all the raw data to construct a long-run time series. For each variable, we impose a priority ordering, based on a careful examination of the quality of the underlying sources. For rates (e.g., inflation, interest rates), we simply combine the raw data in an order of priority.

Variables expressed in levels, as well as some ratios, can exhibit statistical breaks. The GMD thus uses an anchor-year approach to combine different sources. Specifically, we take the data from a reputable source in a recent year as the "truth", and then combine it with other sources using ratio-splicing. The anchor-year fixes historical values preceding a reference year, allowing the dataset to incorporate revisions of recent data without changing historical values. This methodology smoothes jumps that occur when combining different sources, ensuring maximum data consistency over the long-run.

2.5 Archiving and versioning

An important feature of the GMD is that it not only provides regular updates but also an archive of all past versions. This functionality is critical for the reproducibility of empirical results. Because we archive all previous versions and easily make these available through the `gmd` command, this obviates the need for users to store local copies.

3 Introducing the `gmd` Command

The `gmd` command loads macroeconomic data from the Global Macro Database directly into Stata. The command pulls data from the online repository and loads it into memory. In its most basic form, simply typing `gmd` will load the most recent version of the complete dataset with all 74 variables for all countries.

3.1 Syntax

The basic syntax for the `gmd` command is:

`gmd` [*varlist*] [`, version(`*string*`) country(`*string*`) raw vars(`*string*`)`
`sources(`*string*`) cite(`*string*`) print(`*string*`) network(`*string*`) fast(`*string*`)`]

3.2 Options

`version(`*string*`)` loads a specific version of the database. This functionality is critical for reproducibility of empirical results. `current` explicitly calls the most recent version (the default). Versions are specified in the format YYYY_MM. For example, specifying `2025_09` will open the version as of the third quarter of 2025. The command validates the specified version against available releases.

`country(`*string*`)` filters the data to include only the specified country using its ISO3 code (e.g., `USA` for the United States). Typing `gmd, country(USA)` thus opens only U.S. data. `gmd, country(list)` plots the full list of valid country codes. `country(load)` replaces the current data frame with a table showing various country codes.

`raw` opens the underlying raw data processed as part of the GMD, providing all sources for a single variable alongside the GMD estimates. This option requires specifying *exactly one* variable.

`vars` allows users to view all GMD variables. `vars(list)` displays a table with their descriptions. `vars(load)` replaces the current data frame with a table showing variable codes, names, and units.

`sources` allows users to immediately open all (cleaned) data from a given source. `gmd, sources(list)` plots the full list of all available sources. `sources(load)` replaces

the current data frame with a table showing all of the sources. `sources(load)` prints in the console all the available sources.

`cite` is a helpful functionality for generating citations in BibTeX code for a specific source. `gmd, cite(IMF_IFS)`, for example, prints BibTeX code to cite the IMF's International Financial Statistics. `cite(load)` replaces the current data frame with a list of source codes and their BibTeX citation.

`print(`*string*`)` displays the full citation for the GMD citation or the Stata command. For example, `gmd, print(GMD)` displays the citation for the Global Macro Database whereas `gmd, print(Stata)` prints the package paper.

`network(`*string*`)` allows users to bypass the automatic internet check. Specifying `network(yes)` forces the command to assume an active internet connection, which is required for downloading data or checking for updates.

`fast(`*string*`)` allows users to save the data locally instead of downloading each time. Specify `fast(yes)` to do so.

3.3 Basic usage

To load the most recent version of the complete GMD dataset with all variables and countries:

```
. gmd
```

To load the most recent version of the complete GMD dataset with all variables and countries, and save it locally:

```
. gmd, fast(yes)
```

To load specific variables only (e.g., nominal GDP, exports, and population):

```
. gmd nGDP exports pop
```

To load data for a specific country using the `country()` option:

```
. gmd, country(USA)
```

To view the list of available variables:

```
. gmd, vars(list)
```

To view the list of available country codes:

```
. gmd, country(list)
```

To open all data from the IMF's International Financial Statistics used in the GMD:

```
. gmd, sources(IMF_IFS)
```

3.4 Version access

The GMD is updated on a quarterly basis. Each release contains updates to the data, improvements to the documentation, and bug fixes. Due to the open-source nature of the project, we encourage and often receive valuable feedback that we continuously incorporate into the GMD release. By default, the `gmd` command loads the most recent version.

Importantly, the `gmd` command also allows users to directly access all previous releases using the following syntax:

```
. gmd, version(YYYY_MM)
```

To list all previous versions, you can simply type:

```
. gmd, version(list)
```

3.5 Combining multiple options

The `gmd` command allows combining multiple options at once. Loading data on the exchange rate and inflation for Singapore, for example, only requires typing:

```
. gmd infl USDfx, country(SGP)
```

3.6 Access the raw data

The data used as input to the GMD is available for access for all versions:

```
. gmd rGDP, raw
```

3.7 Access data from other providers

Users interested in downloading the data from one specific provider may do so directly via the `gmd` command. Importantly, however, they will only be able to access only the variables that are included in the GMD (which comprises the most widely-used indicators).

To use this functionality, you need to provide the unique code assigned to a data source, which is provided as part of the GMD's technical appendix [4]. The list can also be directly obtained using the `gmd` command:

```
. gmd, sources(list)
```

Once the user knows the source they want to use, they can load the data by specifying it directly in the `sources` option:

```
. gmd, sources(EUS)
```

4. To access the full documentation, please visit: github.com/KMueller-Lab/Global-Macro-Database-Stata.

Note that we only allow loading data from one source at a time.

3.8 Easy citations

The command allows the users to print citations to the data they are using—including the GMD itself—either as APA in-line reference or BibTeX code. For example, to produce a citation for the GMD directly from the Stata command, you can simply type:

```
. gmd, cite(GMD)
```

If you want to cite one of the datasets underlying the GMD, you can easily do that, too. For example, to cite the IMF Historical Debt Database:

```
. gmd, cite(IMF_HDD)
```

3.9 Automatic updates

To ensure that the users always have access to the most recent version of the package, the package automatically notifies users when an update is available.

4 Applications

The strength of the `gmd` command is that it dramatically lowers the "fixed costs" of working with macroeconomic data. No need to manually download datasets, process the data, harmonize units, worry about data irregularities.

To illustrate how easy this makes simple macroeconomic analyses, we use the GMD to investigate two classic empirical phenomena: *Okun's Law*—the correlation between unemployment rates and GDP growth—and the link between money and inflation before and after it was popularized by the monetarist school of thought.

4.1 Revisiting Okun's Law

We begin by studying the link between the unemployment rate and growth in real GDP. The `gmd` command makes pulling a clean dataset with the widest possible coverage of such data trivially easy:

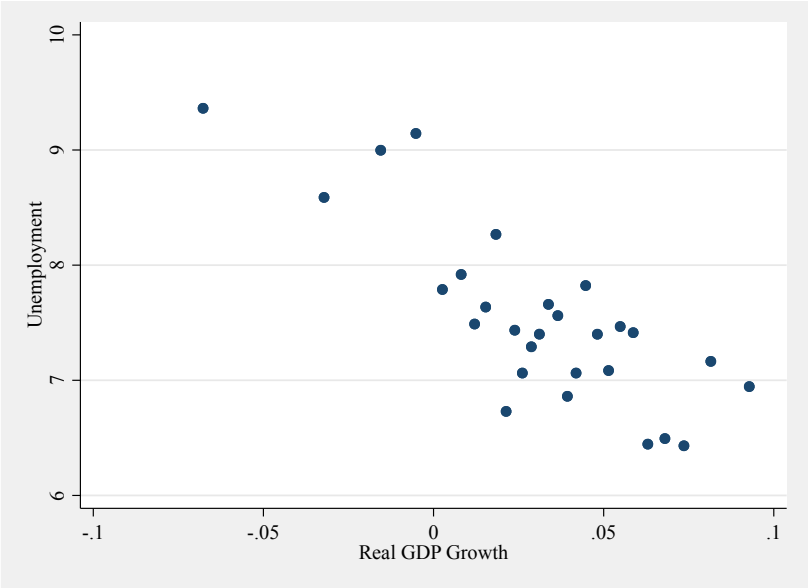
```
. gmd unemp rGDP
```

Next, we need to create a variable capturing year-on-year changes in real GDP growth (here, in local currency units):

```
. xtset id year
. gen rGDP_gr = (rGDP  - L.rGDP) / L.rGDP
```

Equipped with these data on real GDP growth and the unemployment rate, we can create the classic Okun's Law plot. Given the large number of around 7,700 data

Figure 1: Revisiting Okun's Law Using Long-Run Data



Note: This figure plots the relationship between unemployment and real GDP growth using a binned scatter plot, implemented with the Stata package `binsreg`. The data is taken from the Global Macro Database (Müller et al. 2025), obtained using the Stata package `gmd` as specified in the text.

points, we use the user-contributed `binsreg` command for visualization purposes and abstract from extreme values:

```
. ssc install binsreg
. binsreg unemp rGDP_gr if rGDP_gr<0.1 & rGDP_gr>-0.1
```

The figure shows a strong negative correlation. When GDP growth slows down, unemployment spikes, and vice versa.

4.2 Two Eras of Money and Inflation

The GMD can also be used to study the link between money and inflation before and after the release of Milton Friedman and Anna Schwartz's seminal book *A Monetary History of the United States, 1867-1960* (Friedman and Schwartz 1963).

Let's begin by downloading the gmd in its entirety:

```
. gmd
```

As in the previous exercise, we need to create some growth rates. Here, we use M2 as a measure of the money supply:

```
. xtset id year
. gen M2_gr = 100 * ((M2 - L.M2) / L.M2)
```

Perhaps we want to abstract from some very extreme observations, which can happen in the data, and trim both inflation and money growth using the `winsor2` command:

```
. ssc install winsor2
. winsor2 infl M2_gr, cut(1 99) replace trim
```

Finally, we can run two very simple fixed effects panel regressions relating growth in the money supply to inflation rates:

```
. xtreg infl M2_gr if year < 1963, fe cluster(id)
. xtreg infl M2_gr if year >= 1963, fe cluster(id)
```

The coefficient on growth in the money supply (`M2_gr`) is positive and highly statistically significant in both samples. Before 1963, it is 0.48 ($t = 8.33$). Afterwards, it is 0.39 ($t = 9.73$). This exercise highlights just how easy it is to understand relationships between macroeconomic variables using the most comprehensive available dataset using the `gmd` command.

5 Conclusion

In our view, economists are facing an unreasonable burden for working with data that, in principle, should be easily available. The `gmd` command introduced here is designed to help ease that burden by providing users with convenient access to the *Global Macro Database* (Müller et al. 2025), the world's most comprehensive source of macroeconomic statistics.

Apart from making it easy to access ready-to-use macroeconomic data, a few key functionalities are worth highlighting once more. With the `gmd` command, you can access archival versions of the GMD, not just the current version, and you can directly compare the data points from all of the underlying data.

6 References

Friedman, M., and A. J. Schwartz. 1963. *A Monetary History of the United States, 1867–1960*. Princeton: Princeton University Press.

Inklaar, R., H. de Jong, J. Bolt, and J. van Zanden. 2018. Rebasing 'Maddison': New Income Comparisons and the Shape of Long-run Economic Development. GGDC Research Memorandum GD-174, Groningen Growth and Development Centre, University of Groningen. https://ideas.repec.org/p/gro/rugggd/gd-174.html.

Mitchell, B. 2013. *International Historical Statistics*. 1st ed. International Historical Statistics, Palgrave Macmillan London. https://doi.org/10.1057/978-1-137-30568-8.

Müller, K., C. Xu, M. Lehbib, and Z. Chen. 2025. The Global Macro Database: A New International Macroeconomic Dataset. Working Paper 33714, National Bureau of Economic Research. http://www.nber.org/papers/w33714.

About the authors

Mohamed Lehbib is a PhD student in Finance at the National University of Singapore.

Karsten Müller is a Presidential Young Professor and Assistant Professor of Finance at the National University of Singapore. He is also a Research Fellow at the Risk Management Institute and an Affiliate at the CEPR.