

SIPRI's extended military expenditure database 1949-2015 (beta version): an introduction and notes for users

Dr. Samuel Perlo-Freeman, perlo-freeman@sipri.org

SIPRI offers special thanks to Dr. Jennifer Brauner, Elena Deola, Lidwina Gundacker, Dr. Julius Heß, Giulia Tamagni and Mehmet Uye, whose extensive work in searching, documenting, and helping to analyse SIPRI's archived sources of military expenditure data has made this new dataset possible.

25th April 2016

Users should also read the general Sources and Methods for SIPRI's military expenditure database,

http://www.sipri.org/research/armaments/milex/milex_database/copy_of_sources_methods.

SIPRI's published online military expenditure database currently covers the period 1988-2015. Researchers and other users frequently request longer data series that go back before 1988 and now, following several years of data collection and analysis, SIPRI is pleased to be able to make available an extended database, which goes back in some cases as far as 1949. While there are still many gaps in the data series, the database now has data for a majority of countries that existed at the time from the late 1950s onwards.

At present, the extended database is in a 'beta' version, and is being made available on request. SIPRI actively invites users to inform us of any problems or errors discovered in the data, or even better, to make us aware of additional sources of military expenditure and/or economic data that might help us fill some of the gaps. SIPRI intends to publish the final version of the database online, for free access, in September 2016.

Background

SIPRI has been collecting military expenditure data since the late 1960s, publishing it in the SIPRI Yearbook from the very first edition for 1969/70. From the beginning, SIPRI sought to ensure the highest possible level of consistency of data series for each country over time, and after this, compliance with SIPRI's common definition of military expenditure. (See Sources & Methods). When new information and better data series becomes available, past data has been checked and where necessary revised to ensure consistency with the more recent information.

Unfortunately, some methodological changes that were made in the late 1980s and early 1990s, combined with a loss of key records, meant that this historical consistency was lost. After a hiatus in publication, the military expenditure project was relaunched, and a new database created, with a major effort to reconstruct a consistent 10-year series, from 1988-1997, for publication in the 1998 yearbook. Since then, this database has been updated year by year, with consistency back to 1998 maintained. Unfortunately, available

time and resources did not allow project staff at the time, and for many years subsequently, to extend these consistent series backwards before 1988.

The new backdating exercise began in 2010, whereby SIPRI's military expenditure project staff have been assisted by guest researchers—in many cases seeking data for use in their own PhD research—and interns, to sift through SIPRI's extensive archives to collate all available sources of military expenditure data for each country, as far back as possible. These data sources were then compared with each other and with the existing data sources used for the published database from 1988. For each country, we have thus sought to select and combine data series to create the most consistent possible continuous series of data, for as long a time period as possible.

An 'alpha' version of the extended dataset was completed in July 2015, and sent to a group of expert researchers in the field of defence and peace economics and related fields, who as well as making use of the data for research and analysis, helped assess the data for potential errors, and in some cases provided improved sources of data for some countries. The first results of this expert group's research were presented at a workshop held at SIPRI in January 2016, funded by a generous grant from Sweden's Riksbankens Jubileumsfond.

Following this expert feedback, errors have been corrected and some data series revised, leading to the current beta version of the dataset.

Key sources used

A wide variety of sources have been used in reconstructing the extended dataset. The most important of these, by category, are (see also figure 1):

- *Domestic sources of data*, such as government budget and expenditure documents, responses to SIPRI requests for information, national statistical yearbooks, etc. These account for 28% of all data points before 1988. As primary, official data, this is the most preferred sort of data to use, where available. (Although being primary data does not guarantee its accuracy).
- *The IMF Government Finance Statistics (GFS)*, and other IMF publications, accounting for 22% of all data points before 1988. The IMF GFS presents expenditure data reported by governments, using the IMF/World Bank Classification of the Functions of Government, one of which is Defence. The data is not available for all countries and all years, and fewer countries have reported defence data to the GFS in recent years. The COFOG definition of Defence also does not correspond to SIPRI's definition of military expenditure. Nonetheless, for earlier periods it is often one of the best data sources available, and at least represents a common standard.
- *The United Nations Statistical Yearbook (UNSY)*, and the UN Statistical Yearbook for Asia and the Pacific, accounting for 21% of all data points before 1988. UNSY used to provide data on countries' military spending, along with a range of other data. The definition of military spending used is not always clear,

although it very often seems to correspond to the functional definition of 'Defence' spending used by GFS.

- *NATO* data on the military expenditure of its member states accounts for 17% of pre-1988 data, and 55% of the data for European countries. NATO uses a definition of military expenditure very close to SIPRI's.
- *Expert budget analyses and estimates* account for 5% of data points before 1988. Most of these are expert analyses of national budgets and expenditure, and thus are closely based on primary, domestic sources, but often including items of expenditure outside the official defence budget. This category also includes estimates by US analysts for some then-Warsaw Pact nations.
- *Statistisches das Ausland*, a German publication that used to give statistical data on countries worldwide, provides 4% of data points before 1988. This source often, but not always, appears to correspond to the COFOG definition of Defence spending.
- *Other* sources, including the US Arms Control & Disarmament Agency (ACDA), the Europa Yearbook, media sources, other secondary sources, and unidentified sources from previous SIPRI worksheets, account for 3% of the pre-1988 data points.

Methodological issues

The methodologies used for assessing, selecting and combining sources does not differ significantly from that described in the Sources and Methods. However, a few points are worth emphasizing.

It is rarely possible to find a single data source that covers the entire period of available data for a single country. Even developed countries with strong, transparent reporting systems, have frequently changed their methods of reporting and accounting. NATO data provides continuous data back to countries' data of joining, but even NATO have changed their definition, and in some cases have changed the way it is applied to particular countries.

Thus, it has almost always been necessary to combine data from more than one source, and sometimes several, to obtain a long time series of data. To try to ensure consistency over time, a number of key principles were applied:

- Wherever possible, we tried to use data series that *overlapped* in their time coverage, so that the extent to which they agree with each other on overlapping years can be assessed.
- If overlapping series could not be found, we would look for evidence that separate series represented the same *concept* of military spending, for example the Ministry of Defence budget (an institutional definition) or the Defence function based on COFOG (e.g. from the GFS or other sources using the COFOG definition).

- There were, however, a small number of cases where we have combined series without clear evidence of inter-consistency, where this did not give unexpected jumps or dips in real values.
- Where different series did overlap in their time coverage, we would combine the series as is if the difference between them on the transition year was less than about 5%. In some cases if we had several years of overlap, and an inconsistent relation between them (i.e. sometimes one was higher, sometimes the other), we would also combine them directly.
- Where series overlapped, but where one was consistently higher than the other by more than 5%, we would apply a percentage change estimate, adjusting one of the series (usually the older) upwards or downwards by a fixed percentage to bring the two series into agreement on the transition year between them. Such data is marked in blue in the spreadsheets. (This is the same practice as SIPRI has always used for the currently published database, where data is marked blue (in spreadsheets) or in square brackets (in print)). A high proportion of the earliest data is thus marked blue, frequently reflecting the fact that better, more comprehensive data has become available in more recent years, with which older, less comprehensive data, is not consistent.

The need for combining series that may not perfectly agree, and the wide use of estimates, means that there is often considerable uncertainty in the data, although we believe that it accurately captures the broad trends in military spending in the countries covered. Moreover, it should be emphasized that military spending data for many countries, whether recent or older, is often subject to inaccuracy, due to limited transparency in military budgeting and expenditure in many countries. The quality of data may be compromised, among other things, by the use of extra-budgetary or off-budget military spending, poor monitoring and reporting of expenditure, or indeed an absence of data on actual expenditure (as opposed to the initial budget), and other factors. While SIPRI always seeks to identify additional sources of military expenditure outside the main defence budget where possible (for example military spending in other ministries, military pensions funds, or spending on paramilitary forces in police or public security budgets), or at least to identify known gaps in footnotes, there are likely many cases where there are omissions to official data that we do not know about. This is not a new problem to the backdating exercise, but bears restating.

Economic data

Raw data on military spending in local currency at current prices is of limited value if it cannot be compared internationally, or in real terms over time, taking into account inflation in the country in question. SIPRI uses Consumer Price Inflation (CPI) data and market or official exchange rates to convert such raw data into constant price US dollars (usually using a base year of 2014, although for some countries other base years have been chosen due to data availability), and current price US dollars. For more details on how this is done, see the sources and methods. GDP data is also used to express military spending as a proportion of GDP.

Unfortunately, for backdated data, even where military expenditure data has been found for earlier years, it has not always been possible to find the relevant economic data for those years. There are thus a significant number of country-years for which we have data in current price, local currency, but not in constant US dollars.

To minimize this problem, we sought additional sources of economic data, in particular CPI and GDP. SIPRI's primary source for both (in the previous and the extended database) is the IMF International Finance Statistics, supplemented by the IMF World Economic Outlook. Where such data has not been available, the following additional sources have been used:

- Data from the World Bank
- National sources of data (e.g. from national statistical offices)
- Data from older editions of the International Finance Statistics, that are no longer included in the IFS online database
- For inflation data, in some cases GDP deflator data has been used in place of CPI data, taken from either of the IMF sources or the World Bank.

Results of the backdating exercise

The backdating exercise has not succeeded in extending all countries in the SIPRI Military Expenditure Database back to 1949, or alternatively to their date of independence, but has nonetheless produced very encouraging results.

In total, military expenditure series have been extended back to 1975 for 96 of the 130 countries that were independent at the time (87 with constant dollar data, or 67%). Data has been extended back to 1960 for 68 (57 for constant dollars, or 53%) out of 108 countries. From 1957, a majority of countries (46 out of 89) have constant dollar data available. The regions of best data availability are consistently Europe and the Americas. Asia and Oceania is the region with the highest proportion of countries where little or no backdating was possible, but a smaller proportion of African and Middle Eastern countries could be taken back before the mid-1960s. (See figures 2-7).

It should be noted that there are two important countries for which no backdating has been attempted: the Soviet Union and China. Soviet military spending was highly secretive, and while many western attempts were made to estimate it (including by SIPRI), all involved a high degree of uncertainty, in many cases dubious assumptions and methodologies, and sometimes political considerations. In fact, SIPRI stopped publishing estimates of Soviet military spending in the 1980s due to the severity of the difficulties in making even somewhat reliable estimates. Any attempt to produce new and better estimates for Soviet spending before 1988 would require a major new study, and probably an analysis of now-available Soviet-era government documents.

SIPRI's military expenditure estimates for China for 1989 onwards are based on a methodology developed by Professor Wang Shaoguang in 1999, which initially covered the period 1989-98. (See Sources and Methods). It is far from clear that this methodology

would be applicable to the pre-1989 period, and in any case many of the additional data sources used by Professor Wang are not readily available for earlier years. However, the official Chinese defence budget is available back to 1950, although it is likely that the difference between the official budget and the true level of China's military spending (based on a definition close to SIPRI's) was even higher in earlier years. Again, a new study of China would be required to provide meaningful estimates.

‘Crowd-sourcing’ the data

The current version of the extended SIPRI database is a beta version, and is open to continuing improvements. We invite and encourage you, the data users, to help us improve this data, first by pointing out apparent errors, inconsistencies, or improbable changes in the data, and second, by providing us with, or pointing us towards, additional sources of data. For current data sources, SIPRI is heavily reliant on the internet. For the backdated data, covering a period before the internet existed, we have relied solely on sources that were available in SIPRI's archives.

If data for your country (or some other country) is missing for an earlier period, perhaps you know that such data can be found in print sources, such as statistical yearbooks, government reports, etc., that may not have been available to SIPRI when the data was first collected? Maybe a trip to a central library in the capital city can procure the necessary documents. This consideration applies also to the economic data necessary to convert the raw military expenditure data into constant and current US dollars, and to express it as a share of GDP. For example, if data is available for a given country and year in local currency, but not in constant dollars, it is probably because we lack CPI data for that year. Again, perhaps such data is available through print sources in the country in question, even if it is not available online or in standard print reference works such as IFS. Any contributions of data that help us to fill some of the gaps will be gratefully appreciated.

In conclusion, we hope you find this data useful and illuminating, and we hope that, with your help, we will be able to produce an even better version of the dataset for full public release later in 2016!

If you have any questions about this data, please do not hesitate to read the Sources and Methods, at

http://www.sipri.org/research/armaments/milex/milex_database/copy_of_sources_methods, and other material on our website, very carefully to see if you can find the answer. If you don't find the answer on our website, then please do not hesitate to contact us!

Figure 1: Data sources used

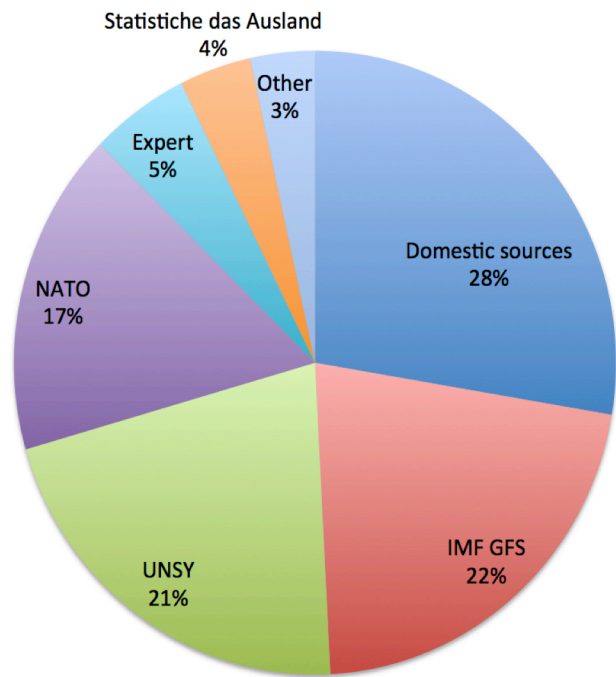


Figure 2: Availability of data, worldwide, 1949-2015

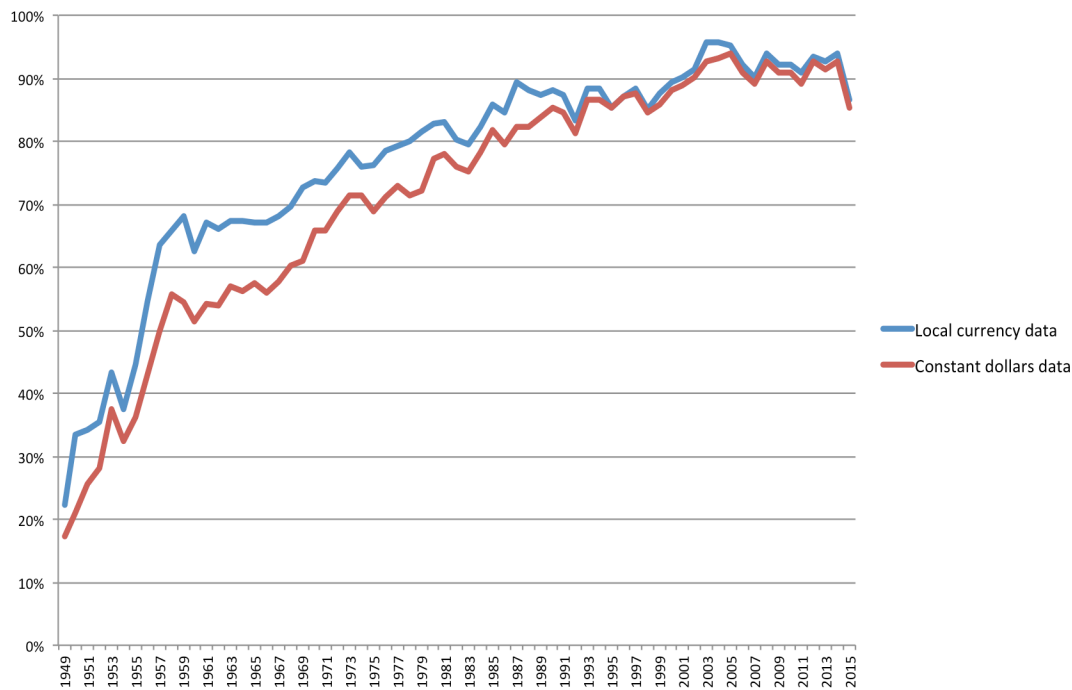


Figure 3: Availability of data, Africa, 1949-2015

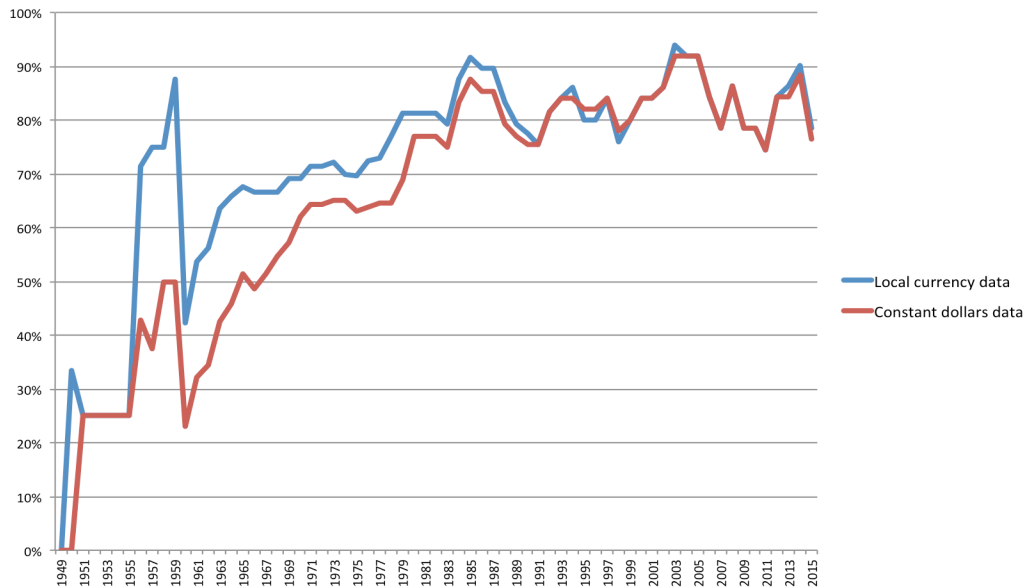


Figure 4: Availability of data, Americas, 1949-2015

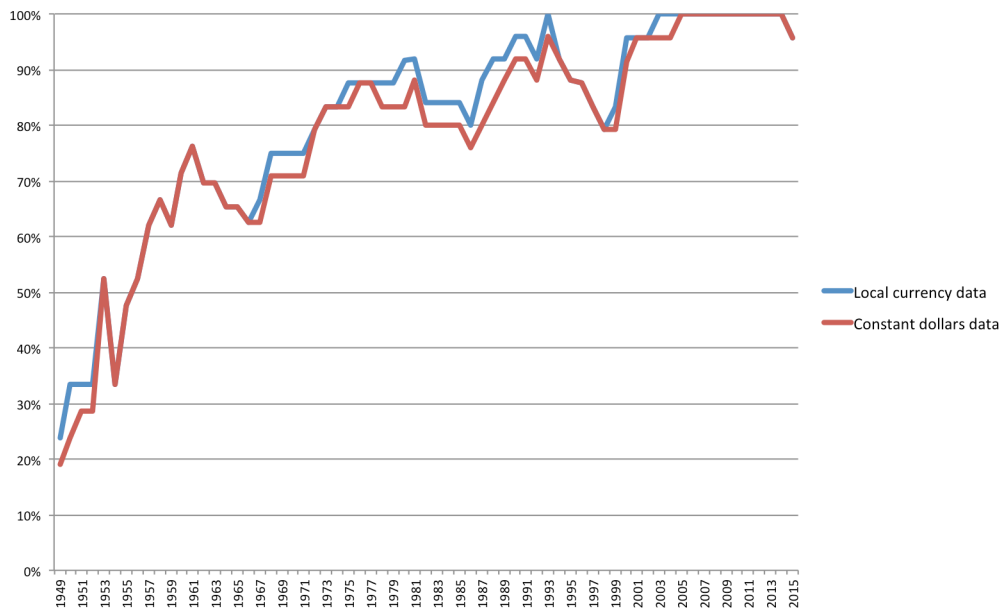


Figure 5: Availability of data, Asia & Oceania, 1949-2015

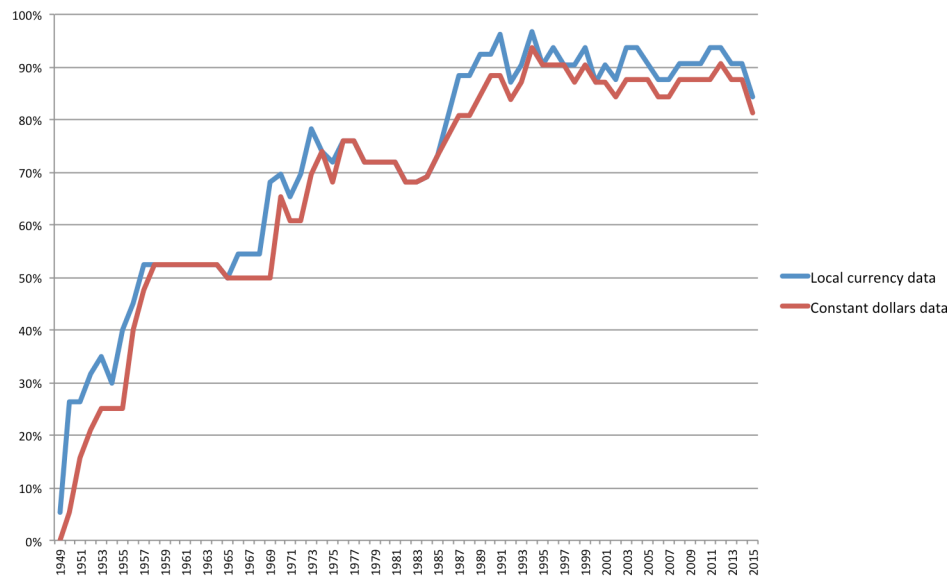


Figure 6: Availability of data, Europe, 1949-2015

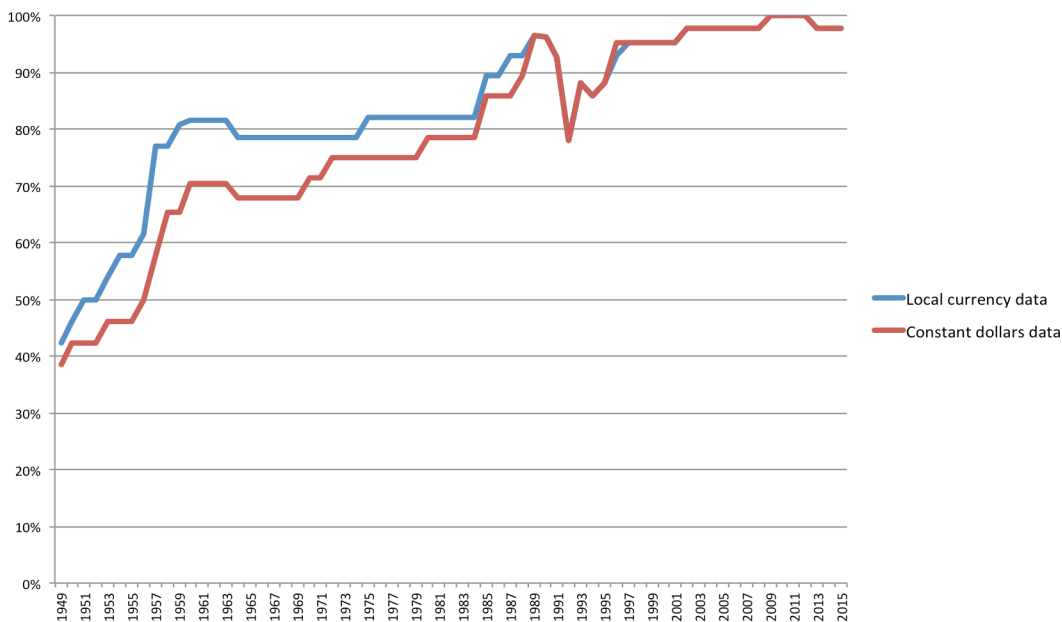


Figure 7: Availability of data, Middle East, 1949-2015

