III. The availability of military expenditure data

NEIL T. N. FERGUSON AND SAM PERLO-FREEMAN

At the core of much of SIPRI's work, and particularly the work to update its Military Expenditure Database, is the question of transparency. In issues of national and international security, transparency remains an issue of great importance, not least because of the opportunity for abuse from authorities that a lack of transparency provides.[1] This section examines patterns in the availability of military expenditure data, based on SIPRI's own data collection efforts, as well as trends in availability over time and patterns between different types of countries.

SIPRI's data collection process brings to light significant variations in transparency between states and, thus, in the quality of data available. In some cases, the available data is detailed and disaggregated into constituent components, such as equipment, operational spending, military pensions, and so on. This allows a deeper understanding of broader trends in the use of state resources and the implications of military spending patterns.

In other cases, defence budgets are presented in much less detail, for example, as broad categories such as 'personnel', 'equipment' or 'operations and maintenance'. Many states provide only a single total figure. Moreover, significant items of military spending—for example, arms purchased through funds from the sale of natural resources, transfers from other departments, 'slush funds' or spending on paramilitary forces—are often excluded from the available data, making the additional spending difficult or impossible to trace.[2]

The extreme case of 'zero transparency' is where no information is available on military spending at all, resulting in missing data points in the Military Expenditure database. Sometimes, the lack of data is no surprise, such as during the recent and current conflicts in Libya and Syria. At other times, however, the reasons are less obvious. Until 2011, for example, data for Viet Nam was available from published government budget documents. Since the defence budget was declared a national secret in 2012, however, only secondary media sources have been available. Despite these issues, there is only one country for which SIPRI has been unable to collect any data since 2000—the highly secretive state of Turkmenistan.

The goal of the analysis presented below, however, is not to attempt to examine or measure the broader transparency issues relating to military

---

[1] For a discussion of the broader issues surrounding transparency in military expenditure see Perlo-Freeman, S. and Solmirano, C., 'Two case studies in the governance of military budgeting and expenditure: Colombia and Indonesia', *SIPRI Yearbook 2013*.

[2] See e.g. Omitoogun, W. and Hutchful, E. (eds), SIPRI, *Budgeting for the Military Sector in Africa: The Processes and Mechanisms of Control* (Oxford University Press: Oxford, 2006).

expenditure, which would require a much more in-depth investigation as well as coding of the available budget data. Instead, this analysis considers the number of years in which there is at least some data available for each country—the absolute minimum level of transparency—in relation to a number of important country characteristics in order to gain a preliminary understanding of the key patterns in the availability of military expenditure data.

The analysis examines trends in data availability by region over time, from 1992 to 2012. It seeks to identify patterns in the types of countries for which data is and is not available, looking at the number of years of data available for each from 2000–11, and relates this to measures of countries' income, level of democracy and state fragility.[3]

Availability of data by region, 1992–2012

Figure 9.4 shows the proportion of countries in each region for which at least some military expenditure data is available in the current database in the period 1992–2012.[4] Western Europe, North America and Oceania are excluded from the analysis as they had a 100 per cent data availability record throughout. The data is based on the new SIPRI dataset up to 2014, but excludes 2013–14 as there is always reduced availability of data in the most recent years. This does not necessarily represent a decline in transparency, but the fact that in some cases data only becomes available at a later date.

Figure 9.4 shows a clear overall increase in data availability from the late 1990s until the mid-2000s and then a decline, but there are clear regional variations (this is represented by the black line in figure 9.4). The most dramatic increase in data availability is in Central and Eastern Europe, the former 'Eastern bloc', which had the lowest figure for any region in 1992 when many of the newly independent states in the region were still highly fragile. Data availability increased dramatically during the 1990s as countries established stable institutions and the wars in the former Yugoslavia ended. It reached 100 per cent in 2002, where it has stayed. Another region that has reached, and maintained, 100 per cent data availability is Latin

---

[3] In 2008 the Organisation for Economic Co-operation and Development (OECD) characterized fragile states as those that are 'unable to meet [their] population's expectations or manage changes in expectations and capacity through the political process'. OECD, *Concepts and Dilemmas of State Building in Fragile Situations* (OECD: Paris, 2008). The World Bank notes that the spectrum of fragility covers 'countries with deteriorating governance, those in prolonged political crisis, post-conflict transition countries and those in gradual but still fragile reform processes'. World Bank, 'Fragile States List', [n.d.], <http://pubdocs.worldbank.org/pubdocs/publicdoc/2014/9/359521410 886172040/FY6toFY9Fragile-States-List-formerly-LICUS.pdf>. Also see chapter 8, section I, in this volume.

[4] Of those that are in the SIPRI Military Expenditure Database and possessed regular armed forces in the year in question.
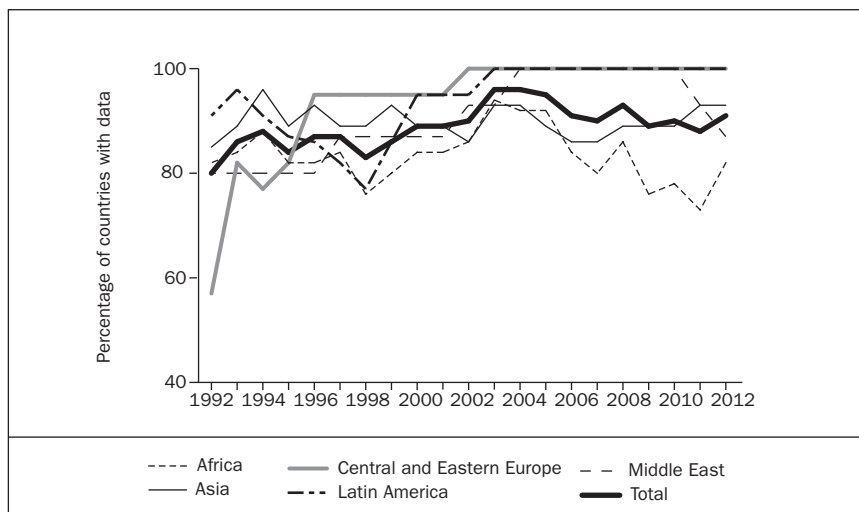
Figure 9.4. Availability of military expenditure data by region, 1992–2012

America, which started at over 90 per cent in 1992, declined to 77 per cent in 1998 but rose to 100 per cent in 2003.

However, the increasing trend in transparency is far from universal. In Asia it reached its peak of 96 per cent in 1994, when at least some data was available for all countries except Afghanistan, but has since fluctuated without a clear trend. Central Asia is the weakest subregion in terms of data availability in 1992–2012. Africa and the Middle East both achieved promising increases in data availability in the late 1990s, but these levels have declined since. In general, data availability in Africa has been weakest in Francophone Africa.[5] In the Middle East, in addition to the lack of data for Syria since 2012, data collection from Qatar has typically been problematic as data is only ever available from secondary sources, such as intermittent reports by the International Monetary Fund.

Data availability by country characteristics

The regional analysis above identifies where there are problems of data availability but gives little sense of why. For example, is the relative lack of data in Africa the result of policy choices on secrecy? Or have weak institutions linked to poverty or fragility led to a failure to pursue effective budgetary procedures or to communicate the information effectively?

---

[5] The SIPRI Milex team researches data sources in English and French.

The sections below present an analysis of data availability, by country, for the period 2000–2011 in relation to the following three economic and political metrics.[6]

1. The World Bank's classification, which includes Organisation for Economic Co-operation and Development (OECD) categories, of countries into five income groups, namely low-income, lower-middle income, upper-middle income, non-OECD upper-income and OECD upper-income.[7]

2. Freedom House's classification of countries as 'free', 'partly free' or 'not free', based on measures of political freedoms and civil liberties.[8]

3. The World Bank's Fragile States List, which classifies countries as 'fragile' if they appeared on this list at least once between 2006 and 2011.[9]

The first two metrics are chosen as potential measures of state *capacity*, that is, whether the country is able to produce and disseminate (in particular online) good quality data in a timely fashion. The third metric is chosen as a variable that may relate to states' *willingness* to provide information, in particular budgetary information, on their militaries to their citizens and to the wider world. Less democratic countries may be more inclined to secrecy in such matters. The initial expectations of the study are that higher-income countries and countries with greater levels of freedom will have higher levels of data availability, while fragile states will have the lowest.

For each of the above measures, and for each category within these classifications, the average number of years in which some data is available was calculated, as well as the proportion of countries where a full set of 12 years of data is available.[10] All the countries in the SIPRI Military Expenditure Database between 2000–11 that have 'non-zero' military expenditure were included in the analysis: 164 countries in total. The results of the analysis are presented in tables 9.5–9.7. They show a clear pattern whereby more data is available for countries in higher-income groups (except non-

---

[6] The SIPRI dataset from 2014 was used for this analysis, with data from 1988–2013, as the data collection process for the 2015 dataset was still ongoing. The past 2 years are excluded because data for the most recent years is often not available until later.

[7] For further information on the World Bank and OECD categories see World Bank, 'World Development Indicators', <data.worldbank.org>.

[8] For further information on Freedom House's classification see Freedom House, 'Freedom in the world 2015: Discarding democracy/return to the iron fist', [n.d.], <https://freedomhouse.org/report/freedom-world/freedom-world-2015#.VTEFla2eDGd>.

[9] For the World Bank's lists of fragile states and situations see <http://pubdocs.worldbank.org/pubdocs/publicdoc/2014/9/359521410886172040/FY6toFY9Fragile-States-List-formerly-LICUS.pdf>; and <http://pubdocs.worldbank.org/pubdocs/publicdoc/2014/9/963681410886171483/FY10toFY13Harmonized-list-Fragile-Situations.pdf>.

[10] Montenegro, South Sudan and Timor Leste were not in existence for the entire period 2000–11. For these countries the number of years of data is scaled up proportionately to be comparable with other countries. Although South Sudan did not become independent until 2011, the military expenditure of the Government of South Sudan, which was created as a result of the Comprehensive Peace Agreement of 2005, is considered from 2006.

Table 9.5. Data availability by income category

Income category	No. of countries	Average years of data, 2000–11	% with complete data
Low-income	32	9.4	41
Lower-middle income	40	11.0	70
Upper-middle income	43	11.4	84
Non-OECD upper-income	18	11.2	83
OECD upper-income	31	11.7	97

Income categories are based on the World Bank World Development Indicators, 2014 classification of countries by income group.

Source: SIPRI Military Expenditure Database, 2014.

Table 9.6. Data availability by state fragility status

Fragility status	No. of countries	Average years of data, 2000–11	% with complete data
Not fragile	132	11.5	86
Fragile	32	8.7	25

Fragility status is based on the World Bank State Fragility Index, 2012.

Source: SIPRI Military Expenditure Database, 2014.

Table 9.7. Data availability by Freedom House civil and political freedom status

Freedom House Status	No. of countries	Average years of data, 2000–11	% with complete data
Free	60	11.7	95
Partly free	55	11.3	73
Not free	49	9.7	51

The Freedom House classification is based on averaging the data from the Freedom House database, 2015, for each country from 2000 to 2011, using the two dimensions of political rights and civil liberties. Countries with an average score of <2.5 are classed as 'free', those with a score of 2.5–5 as 'partly free' and those with a score of >5 as 'not free'.

Source: SIPRI Military Expenditure Database, 2014; Freedom House database, 2015.

OECD upper-income countries) and countries with greater levels of freedom, while fragile states have much lower levels of data availability.

The final section presents the results of a regression analysis, which sought to identify the effects of income, fragility and freedom on the availability of data. Given that, for example, poorer countries are more likely to be fragile, if both these factors appear to be associated with less data availability, regression analysis can help to distinguish whether it is low income, fragility or both that is the driving factor. The results of this analysis show that while freedom and fragility have a clear relation to data

availability, the effect of income is not significant once the other factors are taken into account.

Availability of data by income category

Table 9.5 shows, in general, a clear increase in the number of years of available data as income brackets increase. The exception is for non-OECD upper-income countries, where the mean number of years of available data is between that of the lower- and upper-middle income countries. This tends to reflect the number of oil-rich nations with a tradition of low transparency. When the OECD and non-OECD upper-income countries are combined, however, the mean of the years of available data is greater (11.5) than it is for upper-middle income countries. In general, therefore, income appears to be a strong correlate of transparency when measured in terms of data availability.

An alternative measure to the average number of years of data is to examine the proportion of countries with a complete set of data for each group. This information is also shown in table 9.5, and there is a similar progression through the income groups. Almost all the OECD countries have complete data, but less than half the low-income countries made a full set of data available. The percentage of non-OECD upper-income countries with complete data is almost the same as the percentage of upper-middle income countries.

Fragility

A similar and highly pronounced effect is found in a comparison between fragile and non-fragile states. Only 8.7 years of data is available for countries classed as fragile in at least one year between 2006–11, compared to 11.5 years for non-fragile states (see table 9.6). The importance of fragility is underscored by the fact that among the 15 low-income states that were not classed as fragile between 2006 and 2011, there was an average of 11.3 years of data, while among the 17 low-income fragile countries, the average was just 7.8 years. Thus, without the additional factor of fragility, low income does not appear to be an obstacle to data availability. Table 9.6 shows that 86 per cent of non-fragile states provided a complete set of data, compared to just 25 per cent of fragile states.

Freedom

Freedom House measures the degree of 'freedom' in each country according to two dimensions: political rights and civil liberties. Each is measured using a variety of indicators, leading to a score in each dimension from 1 (most free) to 7 (least free). These two dimensions are averaged to produce an overall score for each country in each year. Countries with an average score of less than 2.5 are classed as 'free', those with a score of 2.5–5 as

'partly free' and those with a score of more than 5 as 'not free'.[11] In this ana-
lysis, the scores were averaged by country over the period in question to
give an overall average score per country of 1 to 7, and the same category
boundaries were then applied.[12]

The results are shown in table 9.7. Countries classified as free have an
almost perfect record, with 11.7 years of data on average. Partly free coun-
tries are only slightly less transparent, with an average of 11.3 years of data,
while not-free countries show significantly less transparency, with just 9.7
years. The pattern is similar when considering the proportion with com-
plete sets of data, but there is a much more marked reduction in availability
for partly free countries—only 73 per cent produced complete data com-
pared to 95 per cent of free countries. To gain a better understanding of
this reduction in data availability, it might be helpful to look at changes in
'freedom' status over time. Some countries where freedom levels have
improved or deteriorated over the period, for example, may have an aver-
age score in the partly free range. It would therefore be interesting to see if
the missing data occurred during the years of lower levels of freedom.

Combining the factors

Such simple statistics belie what are likely to be much more complicated
underlying relationships between fragility, freedom and income. For
example, no low-income or lower-middle income country ranks in the top
20 per cent for freedom and no non upper-income country has the highest
possible freedom score. In this context, the above results raise the question
of whether income or freedom is the key factor behind data availability, or
whether it is a combination of both. In addition: what impact does fragility
have?

One way to attempt to address this question is through simple linear
regression analysis, which allows an examination of the relationship
between several variables at once, and measures the impact of one variable
while holding all other variables constant. Thus, it is possible to test the
impact of income on the availability of data, while taking into account the
effects of a country's freedom rating and fragility, and vice versa.

The results show that a state's freedom ranking has a large, negative and
statistically highly significant effect on data availability.[13] It is, thus,

---

[11] Freedom House (note 8).
[12] Freedom House (note 8). E.g. in 2014 almost all of Western and Central Europe and most of
South America were classified as free, but only a small number of countries in Africa, including
Ghana, South Africa and Tunisia; only India and Mongolia in Asia and Israel in the Middle East were
classified as free. Partly free countries include states such as Colombia, Mexico and Venezuela in
Latin America; Kenya and Nigeria in Africa; Turkey in the Middle East; Ukraine in Europe; and
Indonesia, Malaysia and Pakistan in Asia. China, Russia, most of the Middle East and Central Asia,
and much of Africa were classified as not free.
[13] Full results of the regression analysis and the dataset used to produce them are available on
request from SIPRI.

extremely unlikely that the relationship between freedom and data availability could be the result of chance.[14] In this analysis, rather than using the three broad Freedom House categories as above, the raw freedom scores were used on a scale running from 1 to 7 averaged over the period 2000–11.

The results showed that a marginal decrease in freedom by one unit, that is an increase in the freedom score by one averaged across the period, is associated with a loss of 0.34 years of data. In other words, other factors being equal, a country ranked as having the highest level of freedom (a score of 1) would be expected to have just over 2 (0.34 multiplied by 6) more years of data than a country ranked as least free (a score of 7).

State fragility, too, was found to be a strong and robust determinant of the availability of military expenditure data. The results found that, on average, a fragile state reports 2.32 years of data less than a non-fragile state, given the same income and freedom levels. Once again, this relationship was found to be highly statistically significant, that is, highly unlikely to be purely the result of chance.[15]

Income level measured on a scale of 1 to 5 using the World Bank categories, on the other hand, was found to have a *statistically insignificant* effect on data availability. Once freedom and fragility are controlled for, there is little difference in data availability between income categories, and a high probability that the small differences that are found could be the result of chance.[16]

Given the relative simplicity of these analyses, however, they do not necessarily lead to the conclusion that there is a causal impact running from any of these variables to data availability. Instead, the conclusion is that there are clear correlations that are unlikely to be simply down to chance.[17] Moreover, the factors of fragility, freedom and income cannot be used as bases to fully explain the amount of data available in each country. The regression model above explains 29 per cent of the variation in the number of years of complete data available (the so-called R-squared value).[18] There are, thus, many other country-specific factors affecting the

---

[14] Specifically, there is only a 0.1% likelihood that the result could have occurred by chance.
[15] In this case, there was less than a 0.1% likelihood that the result could have occurred by chance.
[16] The results using 5 income categories found a very slightly *negative* relationship between income group and data availability (–0.11 years per category). Since this may be the result of the non-OECD upper-income countries having lower data availability, analysis was also conducted with the non-OECD upper-income and OECD upper-income categories merged. Once again, the relationship was slightly negative (–0.06 years per category) but statistically insignificant. Other results were unaffected.
[17] The statistical output from these analyses is presented on SIPRI's website to enable full disclosure of the results.
[18] In regression analysis, the R-squared value is a statistical measure of how close the data points are to the graphical line of regression, and thus how good a 'fit' the model is to the data. The R-squared value is equal to the 'explained variation' divided by 'total variation'. The value is always between 0 and 100%, with 0% indicating that the model explains none of the variability of the

outcome. For example, 5 of the 22 countries both labelled not free and classed as fragile at some point between 2000 and 2011 provided complete sets of data, albeit often of a much lower level of detail and transparency than in the best cases. Similarly, each category had countries with at least some gaps in data. Thus, these factors do not fully explain the availability or otherwise of data.

The results suggest that while fragility plays an important role in the availability of military expenditure data, the apparent effects of income are likely to be explained only by their high correlation with state freedom and fragility. This, in turn, raises questions about capacity and transparency—it seems that even the poorest state is just as capable of producing and dis-seminating military expenditure data as an upper-income state with the same fragility and freedom markers. The effects of fragility, on the other hand, are significant, not least as fragile states straddle a number of income strata. In these cases, the institutions that contribute to a state's fragility are also likely to render that country incapable of producing high-quality national accounts, or at least unwilling to do so, particularly on matters concerning the military.

In some countries—particularly oil-producing states, such as Bahrain, Qatar, Russia and Saudi Arabia, which tend to cluster in the non-OECD upper-income stratum—freedom is exceptionally low. Thus, the low number of years of military expenditure data is probably a reflection of choice rather than capability. In fragile states, however, where highly pro-nounced effects were found, it is probably a reflection of poor capacity, rather than a choice to maintain secrecy. That the effects of fragility on transparency are so pronounced suggests a significant danger that fragile states will remain the least transparent, which in turn may lead to further fragility in the future, due to the unchecked vested interests of the elite in continuing to undermine institutional power in such states.

response data around its mean, while 100% indicates that the model explains all of the variability of the response data around its mean.