Total Expenditure on Government Services in the UK, 1980-2000
Data Description and Codebook

April 2004

Stuart N. Soroka
McGill University
Montreal, Quebec
Canada H3A 2T7

Christopher Wlezien
Nuffield College
University of Oxford
Oxford OX1 1NF

Introduction

Here we provide a description and codebook for the Total Expenditure on Government Services in the UK dataset, completed in April 2004. The collection was funded by a grant from the Nuffield Foundation and supported by Her Majesty's Treasury (HMT). The dataset is the product of efforts to create more reliable measures of spending by function. A discussion of problems with available spending series and analyses of the old and new spending data are available in a paper by Soroka and Wlezien (2004) at http://www.degreesofdemocracy.mcgill.ca/measures.

Creating the Dataset

The raw data used as the basis of our collection are from the General Expenditure Statistics team (GES) at HMT. The Public Expenditure Statistics (PES) database is produced and maintained by GES in cooperation with each of the Government Departments. The units of analysis are 'sub-program codes', or SPROGS, the lowest level of aggregation at which spending data are available.

The most recent database years[1] are easily accessible by the GES team. However, older database years, such as the 1985 database, are not so easily accessible. When old databases are removed from the active system, all of the data have been 'dumped into large text-based files. In the past, these files have been unused. There is little documentation provided with them and there is not a specified way—automatic or

[1] A database year is a five-year bloc, containing the current year, the four preceding years and three future projections. For example, database year 2000 would contain data for 1995/1996 through to 2002/2003.

manual—to deal with them.[2] Creating this dataset, then, involved producing final numbers from raw data in a combination of active and in-active database files.

The main reason for making any changes to the data is for consistency. The measure used by HMT has changed over the years, and consequently data that appear in older copies of Public Expenditure Statistical Analyses (PESA) and similar documents are not directly comparable with data in newer copies. More importantly, functional definitions have changed over time. So the same program can move from one function to another from one year to another, and the measured spending change is an unfortunate combination of 'real' change in functional programs, and 'false' change that is the consequence of changing definitions, i.e., rather than actual policy change.

The process of creating consistency is slightly different for each database year, as changes have been made on a continual basis. One of the major problems is that, since the databases were taken off the active system and put into text files, there have been numerous changes to the codes used in the databases. Codes such as Spending Sector, for example, which indicates the sector of government that spent the money, have increased from five numerical codes in the 1985 database, to around ten alpha and numerical codes in the most recent databases. Economic category codes have moved from being just numerical codes in the 1985 database, to alphanumerical codes in the most recent databases.

The consequence of these changes is that there is no simple method for comparing an older database with a current one. Furthermore, the current definition of *Total Expenditure on Services* (TES) relies on these current codes for including and excluding the relevant data, as such; it is essential to determine how the codes compare (using their associated descriptions) in order to determine both what should be included in Total Expenditure on Services, and then the functions into which each program falls.

Data were drawn from the following PES databases:[3]

Database	*Data drawn for fiscal years*
PES85	1980/81 – 1985/86
PES90	1985/86 – 1990/91
PES95	1990/91 – 1995/96
PES00	1995/96 – 1996/97
PES01	1996/97 – 1999/00

As can be seen, from one database to the next there is a year that overlaps, e.g., for PES85 and 90 there is an overlap of the year FY1985/86. Since the last year in each database represents 'work in progress', it is not the most up-to-date data available. In

[2] Information regarding the meaning of codes, such as Spending Sector, Sub-Program, and Economic category were included. However database headings were missing, and no other documentation was present.

[3] Note that forecast years were removed from the databases before they were allowed to leave the Treasury, leaving the databases as four years of outturn data and one year of current data for that year.

these cases, then, it is best to use the data from the newer of the databases. For the example given above, it is appropriate to use data for FY1985/86 from PES90 rather than PES85.

There were four main stages in the process of turning the basic raw data into final functional numbers. All of the processes were carried out in an Excel spreadsheet, and each of the main stages were carried out in individual tabs in the spreadsheet, making it easy to follow the steps logically. These spreadsheets are not included in their original form in the final dataset.

The four steps are as follows:

1. **Total Expenditure on Services definition** – this is the current measure used for public expenditure and our standard. With the exception of the current database, attributes in the database must be manually assessed, to determine which parts should be stripped out.

2. **Functional** – an Excel spreadsheet macro (called the 'lookup sheet'), which is produced and maintain within GES, is run on the data. This attempts to assign each element in the database to a function. However, it does not assign all of the elements, and the propensity for elements not to be assigned increases with the age of the database. This is because differences between the codes used in the lookup sheet and in the older databases increase. Essentially, the lookup sheet is less precise for the older databases, and requires greater diagnosis after being run.

3. **Manual assignation** – the elements not assigned by the lookup sheet must be considered on an individual basis. Each attribute's description and associated codes must be looked at to assess which function the element should be assigned to.[4] This becomes increasingly difficult with the older databases, as the codes used in the particular database become more and more different to current ones, and greater care is required.

4. **Health and Personal Social Services** – these two functions are generally expressed by the Treasury as one function (Health and Personal Social Services in PESA). However, it was desirable for this project to split the two. This was carried out using a slightly different lookup sheet. Similar to the above, some elements will not be assigned to either Health or Personal Social Services, and the manual method of looking at the attributes of the element is required to determine the elements rightful function.

These procedures are not perfect, however. After completing the procedures and collating the data, there were in some cases significant discrepancies between the new

[4] It is possible to assign some elements simply, using the department code, which indicates which department spent the money. With other elements it is necessary to consider detail as low as the sub-program code, and its description. Furthermore, in some cases, it is necessary to combine the information from a number of codes.

series and baseline series created by the GES team using a 'link year method'.[5] Specifically, problems remained in the spending figures for Social Security (SS) and Culture, Media and Sport (CS). In both cases, important component data from earlier years is either missing or unassignable. For SS, a sizeable portion of the spending included in the post-FY1985 data did not appear to be available in the older databases; for CS, some data did not appear to available in the oldest, PES85 database. The resulting series thus show sharp increases in FY1990 (SS) or FY1985 (CS) that are the result of basic database problems and not real spending change. To correct this problem, the series available from the HMT using the 'link year method' were used to interpolate early values for CS and SS. For instance, the percent change from FY1985 to FY1984 in the HMT SS series was used to calculate the FY1984 value for the revised SS series. Although not ideal, this is the best one can do given the available data.

The Final Dataset

The final dataset is available as an Excel file. It includes the new PES series for each expenditure function from 1980 through 2000. The rows designate functions and the columns years. For CS and SS, two sets of numbers are provided: (1) the numbers that resulted from applying our four-stage process; and (2) 'revised' numbers that were the result of the adjustments described above. All of the spending figures are in current Pounds. To provide a basis for comparison, the dataset includes the original PESA numbers from HMT. To aid further analysis, the dataset also includes the Retail Price Index from the ONS, aggregated by fiscal year from the monthly series.

[5] The 'link-year' method combines data from different databases by comparing the difference between the numbers for the overlapping year. The ratio between the estimates using the new and old definitions is applied to the older dataset.