J. R. Statist. Soc. A (2006)
169, *Part* 2, *pp.* 255–271

Public expenditure in the UK: how measures matter

Stuart N. Soroka,

McGill University, Montreal, Canada

Christopher Wlezien

Temple University, Philadelphia, USA

and Iain McLean

Nuffield College, Oxford, UK

[Received April 2004. Final revision August 2005]

Summary. Studying spending over time requires reliable data. It is not clear that such data exist in the UK, however. The two published sources of functional spending numbers—the Office for National Statistics's 'blue book' and Her Majesty's Treasury's *Public Expenditure Statistical Analyses* (PESA)—rely on estimates of past spending, using a link year method, rather than recalculating actual spending figures when functional definitions change. We assess the various measures of spending in the UK. Specifically, we do two things. First, we present a new, third, set of spending numbers applying temporally consistent functional definitions to PESA microdata. Second, we compare the three measures. Our analyses indicate that the Office for National Statistics and PESA data differ quite markedly, especially for certain functions, i.e. in some cases the two measures imply completely different histories. The differences between the original PESA data and our new measures are less pronounced on average, though significant differences are evident, especially year by year.

Keywords: Budget; Fiscal policy; Functions; Policy; Spending

1. Introduction

There are significant problems with all published data on UK general Government spending by function. This news admittedly is not a surprise to many working in the field. Treasury staff, the Office for National Statistics (ONS) and a considerable number of academics agree on the poor state of functional spending estimates (e.g. Heald (1995) and Hogwood (1992)). As John and Margetts (2003) made clear, the lack of quality data has seriously impeded scholarly work in the UK. Nevertheless, little effort has been made to improve the time series—a particularly disheartening situation given the potential importance of these data for policy makers, and for academic analyses of UK policy making and budgetary politics.

Studying spending by function rather than department is particularly worthwhile for those who are interested in using expenditures as a measure of Government policy. The Department for Constitutional Affairs, the Home Office and the Northern Ireland Court Service administer criminal policies and programmes that can be united under a 'law and order' function. An 'environmental services' function can combine expenditures that are managed by the Department for Environment, Food and Rural Affairs, the Office of the Deputy Prime Minister, the Office

Address for correspondence: Christopher Wlezien, Department of Political Science, Gladfelter Hall, Temple University, Philadelphia, PA 19122-6089, USA.
E-mail: Wlezien@temple.edu

 0964–1998/06/169255

of Water Services and various regional governments and agencies. Functions bring similar programmes in different departments together under the same rubric. In effect, they capture the broad aims of policies and programmes; they provide valuable information about the fiscal and policy priorities of governments.

Studying spending over time requires reliable data and, as following sections demonstrate, this has not existed in the UK. The two published sources of functional spending figures present estimates rather than calculations for many years, and the resulting series suggest inaccurate histories. We examine this problem in detail below, assess the accuracy of estimates that are made in the ONS 'blue book' and Her Majesty's Treasury's *Public Expenditure Statistical Analyses* (PESA) and present a new body of spending figures that are based on primary data provided by the Treasury. The new data reflect temporally consistent functional definitions from fiscal year (FY) 1980–1981 to FY 1999–2000, which for the UK Government runs from April 1st of one year to March 31st of the next. A basic comparison of the data indicates that the different measures of spending do differ in important ways. Within both the current blue book and PESA data sets there are questions to be addressed about the accuracy of coding, especially at lower levels in the functional coding hierarchy.

2. The current state of UK functional spending data

There are two sources of UK general Government spending data by function. The ONS blue book is the most commonly used resource. Until 1997, it presented expenditures by function based on general Government expenditure GGE, a measure of Government spending including a number of financial transactions but excluding public corporations. From 1997 onwards, blue book figures are based on expenditure of general Government, a measure which is adjusted to be in line with the 1995 European system of accounts. These post-1997 data are not comparable with the pre-1997 numbers.

The more significant difficulty is that even pre-1997 functional data are unlikely to be directly comparable over a long period of time. Functional definitions change slightly (or not so slightly) each year, so that a programme may be included in health for one year and environment the next. When functional definitions change the ONS does not recalculate most previous years' spending data on the basis of the new definitions. Rather, only the most recent year's $(t-1)$ spending is recalculated. The difference between the original number and the new calculation of spending at $t-1$ is then used to create a factor by which all preceding years' data are multiplied. If, for example, the new calculation of spending on education at $t-1$ is 1.1 times greater than the original estimate, then values for education in all preceding years are multiplied by 1.1 in an effort to bring them in line with the new functional definition. The resulting data thus represent only an estimate of previous spending as it might look by using the new functional definitions. Problems are compounded year by year, so that by 1997 the 1988 data may have been re-estimated nine times.

The other difficulty with blue book data is that, to bring them in line with all other tables in the blue book, and in line with the FY for most other European governments, Government spending is reported by calendar year rather than by UK FY (April 1st–March 31st). Calendar year estimates are not particularly useful for those who are interested in policy making, since spending is allocated and spent by FY. Blue book spending data thus represent an amalgam of decisions that are made in both the current and previous FYs. Since spending will vary considerably by quarter, it is impossible to estimate spending by FY on the basis of calendar estimates.

Thankfully, the Treasury presents functional spending by FY from FY 1980–1981 in PESA. PESA tables are based on a slightly different measure of Government spending—total managed

expenditure, including public corporations but excluding financial transactions (it includes current and capital spending only). If only because spending is aggregated by FY, we should expect the PESA tables to reflect policy making decisions more accurately. This is even more true since 2004, as PESA tables use more detailed and valuable functional categories. Importantly, the PESA figures are generated from the database that is maintained by Her Majesty's Treasury, where the decisions on allocation of expenditure are made. By contrast, the ONS data, when these first appear for a particular year, are the statistical recording of outturns at considerable distance from policy making.

The one difficulty with PESA data is that previous years rely on an estimating procedure which is similar to that used by the ONS. When functional definitions change, data are *recalculated* by using the new definitions for the current year and the preceding four years, because the Treasury maintains 5-year databases. Data previous to year $t-4$ are then *re-estimated* by using the same methodology as described above, referred to by the Treasury as the 'link year method'.

Data gathering methods by the ONS and Treasury reflect their different needs, of course. The ONS priority is statistical record keeping, increasingly (and necessarily) in ways that are in line with European Union requirements. The Treasury is concerned with the operational needs of Government. So it is good that the operational department has a run of FY data for as long as 20 years. The burden of recalculating 20 years of data is too great, however, and so problems of repeated re-estimation remain. And, as a consequence, neither the ONS nor the Treasury data sets provide what researchers may require, i.e. for the empirical testing of theoretical models, researchers may need long FY time series of robust data, not subject to possibly frequent reassignment of expenditure between functions.

So the current state of UK general Government spending data is suspect. Analysts tend to use blue book data—repeatedly re-estimated data that are aggregated by calendar year. The relatively new PESA publication allows researchers to use data that are aggregated by FY, but the problem of using estimates rather than calculations remains. Now, let us see whether and how the measures differ.

3. The data

We have recalculated the figures for every year since FY 1980 by using FY 2000 functional definitions. Using these new data, we can compare the spending series that were published in PESA, *estimated* by the Treasury link year method to match FY 2000 definitions, with series actually *calculated* from the original raw data by using FY 2000 definitions. Results offer an important gauge of the quality of current time series of UK general Government functional expenditures. Moreover, they point to the important differences that reliable measures of functional spending can make.

Our analyses are based on three measures of general Government functional expenditures. First, we examine the GGE data that are drawn from the 1962, 1971, 1980, 1989 and 1997 blue books. These particular blue books were selected by starting with the 1997 blue book—the last before these data were brought in line with PESA. Then blue books were selected at 9-year intervals, as data are given for 10 years in each, and our estimation procedure (which is described below) requires one overlapping year. We do not expect our results to be sensitive to the choice of particular years.

Because functional definitions change and data are repeatedly re-estimated, any given entry for spending on a single function in any given year will differ from one blue book to the next. To avoid introducing shocks into our series that are simply the product of changing blue books, we

employ the same methodology as the ONS and Treasury: data are gathered for an overlapping year from the old and new blue books, factors are calculated and the old series is multiplied by this factor to bring it in line with the new functional definitions. This is not an ideal approach, of course, but we—like the ONS and Treasury before us—have to make do the best that we can.

Second, we examine the Treasury's original series on total expenditure on services (TES) series, TES-O, as presented in Table 3.2 of PESA. TES is just a subset of total managed expenditure, where the latter includes public sector debt interest, net public service pensions, allowance for shortfall and unallocated special reserve, and other accounting adjustments (see Her Majesty's Treasury (2005), chapter 3). These items apply only to the 'total' expenditure series, however; although it is reported in PESA as total managed expenditure by function, functional spending is by definition TES. To be absolutely clear, FY 1995–2000 data for these series are calculated by using FY 2000 definitions; all preceding data are the product of re-estimations to account for changes in functional definitions.

Third, we of course include our new revised spending estimates, TES-R. These estimates use consistent FY 2000 functional definitions for the entire period. They are created by using the 1985, 1990, 1995, 2000 and 2001 5-year Treasury PESA databases, and which contain all expenditures on services at a very low level of aggregation, namely the subprogram code SPROG. In short, we—with considerable help from various people at the Treasury—have recalculated functional spending estimates by aggregating spending by SPROG for FY 1980–1994, using FY 2000 functional definitions. This involved attributing functions to an average of about 4000 SPROGs in each of five PESA databases.

Table 1. SPROGs accounting for 50% of trade and industry functional expenditure in FY 1999†

SPROG	*Description*	*Department(s)*	*Total* ($£ \times 10^3$)
102601	Youth Training Programme	Department for Education and Skills including Office for Standards in Education (OFSTED)	786971
121216	Employment service administration	Department for Work and Pensions	759615
41317	British Nuclear Fuels Ltd	Department of Trade and Industry	477000
40220	Engineering and Physical Sciences Research Council	Department of Trade and Industry	404796
121807	New Deal for Young People	Department for Work and Pensions	352512
121905	European Social Fund payments in advance of receipts	Department for Work and Pensions	312967
151414	Scottish enterprise	Scottish Executive, Scottish Office Education and Industry Department	307989
40221	Medical Research Council	Department of Trade and Industry	304538
40602	Consumer and investor protection	Department of Trade and Industry, Office of Fair Trading, Office of Telecommunications, Office of Water Services, Friendly Societies Registry, Department for Education and Skills including OFSTED	274193
121805	Training for work	Department for Work and Pensions	262309
102101	Departmental administration	Department for Education and Skills including OFSTED	241101
102001	Careers service and employment development projects	Department for Education and Skills including OFSTED	236984

†Source: Her Majesty's Treasury, PESA Team, 2001 general expenditure statistics database.

Table 2. Largest subprogrammes in the transport function†

SPROG	*Description*	*Expenditures (£ × 10^6) for the following FYs:*			
		1998–1999	*1999–2000*	*2000–2001*	*2001–2002*
060101	Highways Agency programme	1355.4	1391.1	1413.1	1548.3
060325	Strategic Rail Authority	—	—	147.8	1229.8
060311	Office of Passenger Rail Financing (now within Strategic Rail Authority)	1024.2	890.4	603.8	—
060511/09	London Underground	241.0	178.8	—	483.0
060507	London Regional Transport	166.5	156.2	315.4	—
061005	Channel Tunnel	115.9	11.5	27.3	358.3
060307	Bus fuel grants	225.0	292.4	299.6	303.9
060779	Central administration	113.0	112.0	118.9	102.0
Total of these subprogrammes		3241.1	3020.9	2898.6	4025.3
Total spending		3904.0	3843.9	3486.4	4610.1

†Source: Department for Transport, Local Government and the Regions (now the Department for Transport) returns to Her Majesty's Treasury PESA database, collected and reported by McLean *et al*. (2003), Table 4.3.1.

This was done in a two-stage process. First, functional categories were assigned by using a 'lookup sheet'—an EXCEL spreadsheet macro which seeks to attribute functions to each SPROG automatically. This sheet is maintained by the Treasury. We used the FY 2000 lookup sheet, which predictably has a very good success rate in recent years but cannot assign an increasing number of SPROGs as we move back to FY 1980. To be specific, the number of elements that were unassigned by the macro was negligible in recent databases—none in the 2000 and 2001 databases, and just 32 in the 1995 database (out of 3430 final, functionally assigned SPROGs). In early databases, however, the number of elements that were left unassigned was considerable—about 1100 (of 4935) in the 1990 database and 1400 (of 3440) in the 1985 database.

The second stage of the process involved assigning functions to these remaining elements by hand. The difficulty with which this was done varies considerably. In some cases, it was possible to assign functions by using the department code, which indicates which department spent the money. In other cases, we needed to consider the brief description of the SPROG that was included in the database; sometimes this could be compared with successfully assigned SPROGs in subsequent years, and a functional category assigned in that way. In a limited number of cases, it was necessary to combine the information from a number of codes to assign a function. There is necessarily some subjectivity in these decisions, then, though no more than there is in any effort to assign spending to functions, either historically or currently. The important point here is that the bases for assigning functions to SPROGs remained constant across the entire period—i.e. we worked with a single set of functional definitions. We benefited enormously here from working with the Treasury staff—indeed, work on the 1985 database was done over the course of 2002 in the Treasury offices themselves.

Tables 1 and 2 provide some further details about the information in the PESA databases, and both the difficulty and the importance of revising current ONS and Treasury estimates. Table 1 shows the difficulty of re-estimating functional spending, using trade and industry as an example. Table 1 shows, in descending order of size, the largest SPROGs in the trade and industry function, which account for over 50% of the expenditure in this domain. This list includes most

Table 3. Revised measures of TES by function

Function	*Expenditures (£ × 10^9) for the following FYs:*									
	1980–1981	*1981–1982*	*1982–1983*	*1983–1984*	*1984–1985*	*1985–1986*	*1986–1987*	*1987–1988*	*1988–1989*	*1989–1990*
Education	12.882	14.009	15.053	15.849	16.412	17.541	19.227	21.007	22.684	24.997
Health	12.365	13.878	15.099	15.984	17.108	17.634	18.869	20.699	22.912	24.744
Personal social services	1.702	1.876	2.048	2.221	2.367	3.050	3.351	3.765	4.197	4.713
Transport	4.282	4.842	5.391	5.623	5.738	5.578	5.450	5.668	5.516	6.486
Housing	5.140	3.446	3.175	4.747	4.817	4.700	4.708	5.039	5.116	6.058
Other environmental services	3.307	3.417	3.835	4.167	4.340	3.832	3.932	3.864	2.965	4.502
Law and order and protective services	3.886	4.536	5.031	5.543	6.096	6.386	6.925	7.685	8.530	9.739
Defence	11.203	12.641	14.441	15.534	17.161	17.933	18.185	18.856	19.079	20.760
International development assistance and other international services	1.159	1.168	1.230	1.336	1.484	1.659	1.777	1.827	2.048	2.282
Trade, industry, energy and employment	4.598	5.409	6.039	6.284	7.840	6.984	7.004	6.421	6.700	6.415
Agriculture, fisheries, food and forestry	1.642	1.675	2.171	2.449	2.439	2.781	2.111	2.318	2.098	2.051
Culture, media and sport	1.160	1.396	1.525	1.557	1.668	1.787	1.906	2.144	2.382	2.740
Social security	23.518	28.732	32.892	36.868	39.673	43.270	46.571	48.643	49.761	52.754
Central administration and associated expenditure	2.252	2.712	2.875	2.810	3.340	5.018	5.982	6.575	6.163	8.560
TES	89.096	99.739	110.804	120.971	130.482	138.153	145.998	154.510	160.151	176.803

(*continued*)

Table 3 (*continued*)

Function	*Expenditures (£ × 10^9) for the following FYs:*									
	1990–1991	*1991–1992*	*1992–1993*	*1993–1994*	*1994–1995*	*1995–1996*	*1996–1997*	*1997–1998*	*1998–1999*	*1999–2000*
Education	26.591	29.181	31.505	33.029	34.604	35.573	36.143	37.181	38.766	40.680
Health	27.693	31.639	34.989	36.680	38.984	41.694	43.019	44.954	47.528	51.438
Personal social services	5.399	5.916	6.388	7.248	8.508	9.428	10.110	10.691	11.352	12.505
Transport	8.895	9.772	11.367	10.472	12.000	11.508	10.076	9.231	8.698	8.505
Housing	5.333	6.109	6.393	5.657	5.133	5.029	4.579	3.718	3.679	2.765
Other environmental services	7.241	7.653	7.767	7.669	8.727	8.498	8.639	8.817	8.712	10.054
Law and order and protective services	11.314	12.755	13.949	14.502	15.127	15.737	16.375	17.007	17.475	18.760
Defence	21.709	22.913	22.910	22.622	22.522	21.631	21.324	20.946	22.634	22.717
International development assistance and other international services	2.531	2.924	3.137	3.260	3.269	3.303	2.952	2.869	3.165	3.429
Trade, industry, energy and employment	10.412	9.670	10.204	10.584	9.789	9.189	8.900	8.696	8.951	9.337
Agriculture, fisheries, food and forestry	2.919	3.058	3.182	4.089	3.657	4.170	5.409	4.686	4.647	4.408
Culture, media and sport	2.978	3.035	3.137	3.122	3.281	3.438	3.653	4.162	4.926	5.565
Social security	58.843	69.954	79.264	86.597	88.943	92.754	96.436	97.365	99.403	103.222
Central administration and associated expenditure	6.906	6.044	8.324	8.279	7.671	9.833	8.090	8.277	10.414	10.272
TES	198.765	220.623	242.517	253.811	262.214	271.784	275.705	278.600	290.349	303.656

of the departments whose PESA returns were aggregated in the trade and industry function (see also McLean *et al.* (2003), paragraph 4.8.1). Table 1 testifies, then, to the significant differences between departmental and functional spending. Moreover, it shows just how heroic an assignment of all SPROGs to an appropriate function of Government is. Unless immersed in the data, the analyst may underestimate—or simply misestimate—the variety of public expenditure in a modern industrial state.

In short, the taxonomy of functions of Government is not like the taxonomy of naturally occurring organisms, which Dawkins (1986), page 259, labelled 'perfectly nested'. He went on:

> 'This property of perfect taxonomic nesting is not exhibited by books, languages, soil types, or schools of thought in philosophy. If a librarian draws a ring round the biology books and another ring round the theology books, he will find that the two rings overlap. In the zone of overlap are books with titles like 'Biology and Christian Belief.'

So it is with functions of Government. In the UK, the Ministry of Defence is responsible for the defence of the realm, and the Department for Work and Pensions for the payment of social security benefits. So who pays social security benefits to former defenders of the realm and their relicts—i.e. war pensions? It could be either department, but the function must be assigned to just one. And war pensions have been moved from one to another function—the relevant SPROG has been reclassified—over time.

Thus, even pairs of functions that seem remote from one another have a common frontier across which SPROGS are shifted from time to time. The result is that apparent changes in spending are sometimes only partly related to changes in spending, and partly to changes in functional definitions. Spending on SPROGs can move around extensively every year, after all. Table 2 shows how much this can happen within a single functional category.

Table 2 shows the movement in the top eight transport SPROGs in each of the four FYs from 1998 to 2001. These eight SPROGs account for about 80% of UK public expenditure for this function. During the period in question, the Strategic Rail Authority was created. It has now been abolished, and by the same administration that introduced it. A substantial proportion of London transport expenditure was moved from London Underground to London Regional Transport, and then back again. Expenditure on the Channel Tunnel rail link fluctuated hugely. The first two of these sets of movements are bookkeeping changes; the last is (probably) a real change in public expenditure. The aim of our research was to separate these two sets of changes. Our measure TES-R assigns all SPROGs back to FY 1980 to the function to which they were assigned in FY 2000. TES-R thus allows the analyst, as TES-O does not, to isolate the *real* changes in public expenditure, such as those on the Channel Tunnel, by assigning these sub-programmes to the same function of Government for each year since FY 1980. The assignment is necessarily partly arbitrary, but it is *consistent* over time. This is why we see an advantage to TES-R over TES-O and GGE.

The new data resulting from our re-estimation are contained in Table 3. Note that these data necessarily end in FY 1999. For years before FY 2000, UK Government spending was controlled on a cash basis, where payments and receipts were recorded in the year in which they actually occurred. The 2000 spending review marked the beginning of resource budgeting, however, where resources are recorded as they are consumed (rather than paid for). A concise description of the difference between cash and resource budgeting is provided in Her Majesty's Treasury (2000). For our purposes, the move to resource budgeting means that data that are gathered from FY 2000 onwards are not directly comparable with preceding years' data.

Table 4. Descriptive statistics for different budgetary measures, 1980–1996†

Function	*Results (1987 £ × 10^6) for the following measures:*					
	GGE		*TES-O*		*TES-R*	
Agriculture	2498	(446)	2537	(481)	2486	(430)
Central administration	—	—	5604	(707)	5099	(1260)
Culture, media and sport	—	—	1980	(256)	2081	(229)
Defence	17736	(1255)	16882	(1491)	16951	(1454)
Education	21255	(2478)	20341	(2361)	20679	(2137)
Health and personal social services	—	—	25343	(4615)	25771	(4962)
Health only	21881	(3706)	21224	(3441)	21786	(3638)
Personal social services only	—	—	4119	(1205)	3984	(1351)
Housing	8090	(1606)	4383	(1178)	4602	(1019)
International development	—	—	1852	(222)	1888	(271)
Law and order	8037	(1760)	8293	(1773)	8131	(1789)
Other environmental services	—	—	4921	(492)	4792	(904)
Social security and personal social services	53614	(9964)	—	—	—	—
Social security	—	—	48725	(8973)	48838	(8939)
Trade and industry	5527	(2800)	7230	(976)	7031	(1007)
Transport	5182	(1059)	7388	(728)	6567	(943)
Total for functions listed	143820		180822		180686	

†Entries are mean levels of spending; numbers in parentheses are standard deviations.

4. Comparing new and old measures

Having described the data, we now move on to some preliminary analyses. All analyses include those years for which data are available for all three data series: 1980–1996. All spending series are also transformed into constant 1987 British pounds by using a gross domestic product deflator (1987 ≡ 1.00). A calendar year deflator for GGE estimates and an FY deflator for TES estimates were calculated from the monthly series (CHAW and CBAB, which are available from the ONS).

Table 4 presents basic descriptive statistics for the GGE, TES-O and new TES-R series by all available budget functions. Recall that TES-R series start in FY 1980, and GGE series end in 1996, which limits our analyses to 1980–1996. Also recall that yearly data for GGE are necessarily by calendar year (1985 ≡ January 1985–December 1985), whereas TES-O and TES-R data are by FY (1985 ≡ April 1985–March 1986). Also, functions that are listed in the blue book are not identical to those that are listed in PESA, so data are missing in some of the cells.

In Table 4 we can see some significant differences between the spending series. Housing estimates, for instance, are much higher for GGE than for TES-O or TES-R; trade and industry estimates are lower. Differences between ONS and Treasury functional definitions are a likely culprit, but even the TES-O and TES-R series show differences in some domains, including culture, media and sport, housing, trade and industry, transport and international development. Significant differences do not exist between the three spending estimates across all domains, however, and this is a *leitmotiv* for our examination: although there are important differences between the various spending series, the magnitude and direction of these differences vary considerably across functions. The difference between TES-O and TES-R for defence tells us nothing about the differences between these series for education.

Table 5. Correlations between budgetary measures, in levels and differences, 1980–1996†

Function	*Correlations for the following comparisons:*					
	GGE–TES-O		*GGE–TES-R*		*TES-O–TES-R*	
	Levels	*Differences*	*Levels*	*Differences*	*Levels*	*Differences*
Agriculture	0.869	0.867	0.895	0.863	0.975	0.966
Central administration	—	—	—	—	0.676	0.740
Culture, media and sport	—	—	—	—	0.980	0.942
Defence	0.967	0.664	0.964	0.596	0.997	0.980
Education	0.987	0.810	0.984	0.759	0.994	0.920
Health and personal social services	—	—	—	—	0.998	0.918
Health only	0.996	0.685	0.993	0.647	0.992	0.919
Personal social services only	—	—	—	—	0.991	0.647
Housing	0.855	0.638	0.786	0.725	0.844	0.863
International development	—	—	—	—	0.950	0.733
Law and order	0.996	0.821	0.997	0.861	0.996	0.889
Other environmental services	—	—	—	—	0.730	0.442
Social security and personal social services	—	—	—	—	—	—
Social security	—	—	—	—	1.000	0.999
Trade and industry	0.521	0.119	0.282	0.195	0.469	0.346
Transport	0.031	0.016	0.186	0.166	0.896	0.889
Mean	0.778	0.578	0.775	0.566	0.877	0.675
Median	0.918	0.725	0.930	0.686	0.980	0.889

†Entries are Pearson correlation coefficients. $N = 17$ for levels, and $N = 16$ for differences.

Consider Table 5, which shows pairwise Pearson correlations between the three series for 1980–1996. Correlations are shown for both levels and differences. For differences, the value for a series at t is equal to the level of spending in year t minus the level of spending in year $t-1$; these correlations accordingly capture the extent to which *changes* in two series are the same. Predictably, the correlations tend to be lower in differences than in levels, though this is not always true, such as for the correlation between the two TES measures for central administration. The mean and median correlations are substantially higher in levels, however.

Three patterns are worth noting. First, evidence again suggests that there are considerable differences across domains. GGE, TES-O and TES-R series closely track each other in some areas, but not in others. The education, and law and order functions display relatively high correlations in both levels and differences; transport correlations are low, at least when including the GGE series; trade and industry correlations are consistently low.

Second and perhaps most striking are the comparatively low correlations between the GGE measures on the one hand and the two TES series on the other. This is understandable, given that both the functional definition and the period of aggregation are different for GGE from those for TES. Still, mean correlations of only about 0.78 in levels and 0.57 in changes between GGE and either TES series point to some quite significant differences between the spending measures. In particular, these differences raise serious questions about the value of aggregating spending by calendar year rather than FY. European Union regulations aside, it appears that GGE estimates simply *cannot* provide accurate indications of UK budgetary policy.

Finally, there are potentially significant differences between TES-O and TES-R measures in certain areas, particularly housing, international development, other environmental services, and trade and industry. The lack of congruence between these two TES series is probably the product of two on-going phenomena:

(a) some programmes are particularly difficult to categorize by function and
(b) for some functions, definitions change more frequently over time.

Even for seemingly straightforward functions like defence and health, the different measures produce slightly different estimates. But the low correlation in differences for other environmental services is a signal that particular SPROGs in this diffuse group have been frequently moved around from one main function of Government to another. For corroboration, note that the UK Government department with Environment in its title has been sometimes the department whose core function is managing central–local government relationships (Department of the Environment in the 1970s; Department for the Environment, Transport and the Regions in the 1990s), and at other times the department whose core function is managing the countryside (the Department for Environment, Food and Rural Affairs since 2002). That TES-R carefully controls for such transfers of function makes it a better analytic series than TES-O, which does not.

That differences exist between TES-O and TES-R confirms the potential importance of recalculating historical data by using current functional definitions. Even so, it is clear that this matters more in some domains than in others. Some of this variation is systematic—functions in which spending is less incremental, in which yearly changes tend to be proportionally greater, such as trade and industry, tend to change more from recalculation than do functions in which spending is very highly autocorrelated, such as education. This is predictable, since the Treasury's link year method—described above—will tend to minimize what Jones *et al.* (1998) defined as 'punctuations' in spending. (See a further discussion of this later.)

Many of the differences between the various spending series are unpredictable, however—i.e. there is no simple story about how one set of spending estimates differs from the other. The various potential differences in measures are illustrated in Fig. 1, which plots GGE, TES-O and TES-R for three functions—defence, transportation, and trade and industry—during the period between FY 1980 and FY 1996. The functions are selected because they show the general ways in which the series might differ. In some cases, such as defence, the differences in the measures are mostly in levels, and these can be quite sizable. The gap between the GGE estimates and the two TES series is virtually 10% of the TES-O (or TES-R) defence budget. Things are much the same for transportation, though here we can detect level differences across all three measures. We also can detect significant differences in year-to-year changes over time, at least between the GGE series on the one hand and the two TES series on the other. For trade and industry, differences exist in both levels and changes throughout for all three measures. The history of spending on this function depends almost entirely on which measure is used.

Some of the differences between measures must be due to varying periods of aggregation—GGE being aggregated by calendar year rather than FY—or differences in the functional definitions that are used by the ONS and the Treasury. In many cases, however, the differences between series are by and large the product of repetitive re-estimates of spending by function, precipitated by over-time (and unrecorded) changes in functional definitions. For budgetary scholars, simply identifying these differences is an important task. For others, the importance of this work relies in large part on demonstrating that the differences in spending series have real effects on substantive analysis. We address this below.

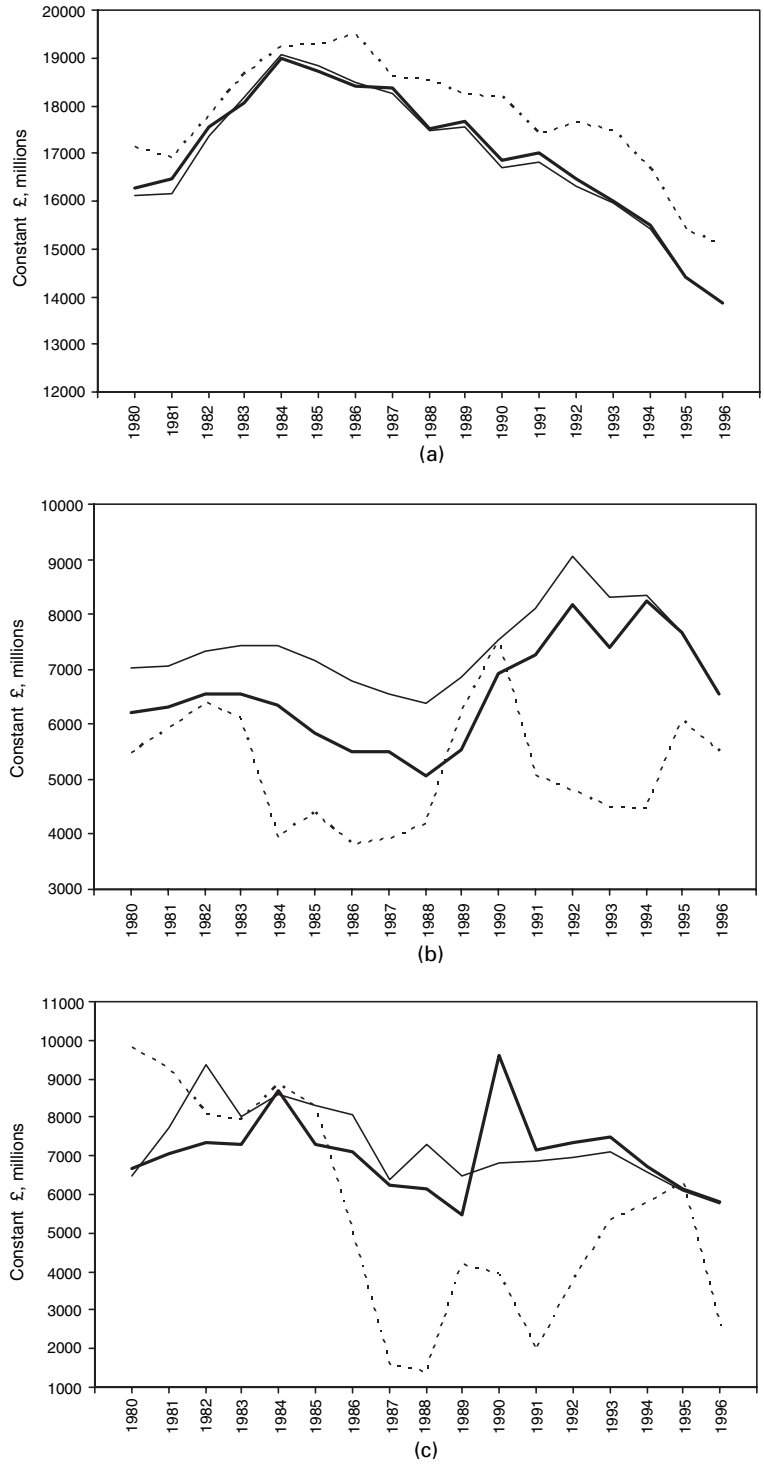

Fig. 1. Comparing spending measures (-------, GGE; ———, TES-O; ———, TES-R): (a) defence; (b) transportation; (c) trade and industry

5. An expository analysis: policy punctuations

Fig. 1 suggests that our re-estimates of functional spending may make significant differences to more developed quantitative analyses of Government spending. To make this point more clearly, however, we draw here on the recent literature on punctuations in Government spending. Baumgartner and Jones's (1993) punctuated equilibrium model of policy change has received increasing attention in recent years, both in the USA and elsewhere. The model posits that policy will tend to move incrementally most of the time, owing to a combination of institutional constraints and the bounded rationality of individual decision-making. (For more complete discussions of the incrementalist model of policy making, see Lindblom (1975), Wildavsky (1984) and Jones and Baumgartner (1997).) Punctuations in policy also will occur, though less frequently, for instance because of the change in the party control of Government or else substantial shifts in the perceived need for spending (see, for example, Wlezien (2004)). Ultimately, then, policy should be characterized by a great number of small changes and a few very large changes—long periods of equilibrium with occasional but considerable punctuations.

Jones *et al.* (2003) found that a wide variety of political time series—budgetary and otherwise—are characterized by long periods of stasis interspersed with dramatic punctuations. Moreover, the extent to which political series fit the punctuated equilibrium model increases as 'costs' increase—systems in which information is difficult to obtain and where there are considerable institutional constraints on decision-making, for example, fit the punctuated equilibrium model particularly well. Public budgeting is one such system, as evidenced by the distribution of percentage changes in spending over time. A histogram of percentage changes in US functional spending reveals a *leptokurtotic* distribution: a non-normal distribution that is characterized by a slender peak (the product of incremental change) and heavy tails (the product of infrequently large change).

John and Margetts (2003) recently tested Jones's hypothesis in the UK by using the GGE database that we compiled for the present study. They compared the distribution of changes in spending across a number of domains from 1951 to 1996, exploring variations in leptokurtosis, and found ample evidence of punctuated equilibria in UK functional spending. How would such an analysis differ by using the three spending measures that we use here? Fig. 2 directly compares distributions for GGE, TES-O and TES-R, from 1978 to 1996. It relies on percentage changes—as did the analyses of Jones *et al.* (2003) and John and Margetts (2003)—in the three spending series (in constant 1987 pounds), for all eight functions for which all data are available. A normal curve is shown in each graph, as are summary statistics.

Fig. 2 indicates that the distributions of percentage changes in all three series indeed exhibit abnormally high peaks and long tails, in line with work by Jones *et al.* (2003) and John and Margetts (2003). The pattern in Fig. 2 is supported by estimates of Pearson's k, which provide an indication of kurtosis (see Anscombe and Glynn (1983)) and reveal positive kurtosis in all three distributions. This is not the ideal test, as summing different normal distributions—even with the same means—but with different variances can produce leptokurtosis. There are in this case differences between the three measures, however. GGE exhibits the most leptokurtotic distribution, i.e. it shows the greatest proportion of large punctuations, well above and below the zero point. (For clarity of illustration, the graph—though not the k-statistic—for GGE estimates excludes a single point at 203.6% change for trade and industry in FY 1989.) By contrast, the TES-O series shows a higher peak at about 0, indicating a considerable amount of incremental change, and relatively few large punctuations. TES-R is somewhere in between, as illustrated by both the graph and the k-statistic.

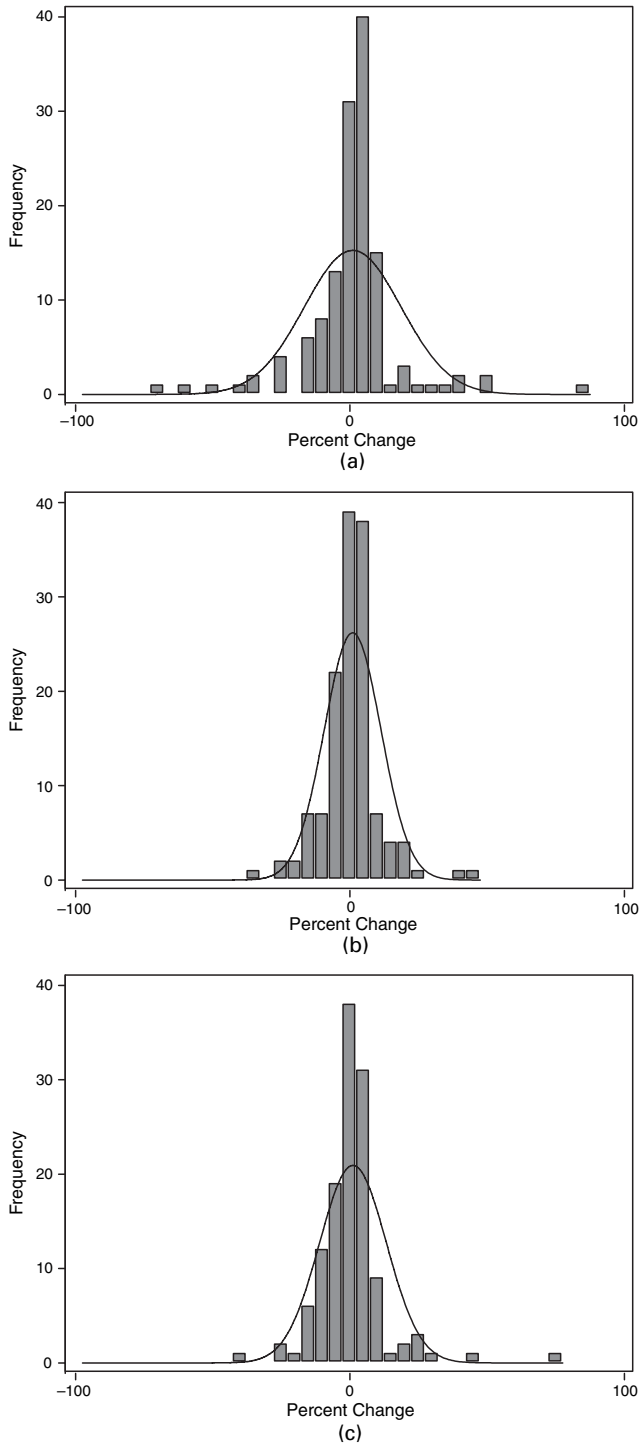

Fig. 2. Comparing spending measures—policy punctuations: (a) GGE (mean, 2.49; standard deviation, 24.72; $k = 35.3$); (b) TES-O (mean, 0.98; standard deviation, 10.35; $k = 7.7$); (c) TES-R (mean, 1.09; standard deviation, 12.20; $k = 14.5$)

Table 6. Numerical assessment of the tails†

Tail	*Results for the following spending measures:*		
	GE	*TES-O*	*TES-R*
Right	2.39	1.59	1.82
Left	1.59	1.38	1.40

†The statistics represent the root-mean-squared change in spending divided by the mean change in spending.

The same is true when we look more closely at the shape of the tails in each distribution. This can be done in various ways (see Jones *et al.* (2003)). We rely on the ratio of the root-mean-squared change in spending to the mean change in spending. This ratio taps the length of the tails relative to a normal distribution of the same variance; it is 0.80 for Gaussian tails, i.e. a normal distribution, and a larger ratio indicates longer, thicker, tails, much like the k-statistic. The ratio can be calculated separately for the positive and negative tails of the distribution for each of the three measures. Results of doing so are presented in Table 6, and these generally confirm what we saw in Fig. 2. First, for each measure, the ratio for the positive section is larger than that for the negative section. This indicates longer, thicker, right-hand tails. Second, for each tail, the ratio for GGE is larger than for both TES measures. This implies thicker tails for the GGE distribution, though especially in the positive section, as in Fig. 2. Third, the ratios for TES differ only for the right-hand tails, where TES-R shows less scatter than GGE but slightly more than TES-O. The difference is not fundamental, however.

These results are perhaps what we should expect. For GGE, it may be that a combination of

(a) changes in functional definition and
(b) aggregating the final quarter from the previous FY and the first three quarters of the current FY

enhances—or, rather, creates—punctuations in spending estimates. (The punctuations in GGE clearly cannot be the simple consequence of putting together data from different blue books, as the backward smoothing method that we use to connect data over time minimizes punctuations as we move from one source of data to another.) In contrast, TES-O aggregates by FY, but relies on a large amount of backward estimating. This necessarily smooths out fluctuations in the data. TES-R resolves both problems—the data are aggregated by FY and rely on temporally consistent calculations rather than backward estimates. The TES-R data consequently offer a more accurate picture of UK fiscal policy making.

6. Discussion and conclusions

In spite of the potential value for both policy makers and policy analysts, there is no published reliable source of UK GGE by function over time. Practitioners and scholars alike rely on two sets of time series: the GGE series from the blue book and the total managed expenditure series from the Treasury. What we have shown is that these series differ in meaningful ways. The differences may in large part be the product of varying periods of aggregation—specifically, by calendar

year instead of FY. Continuously changing functional definitions and repeatedly re-estimated rather than recalculated spending figures also appear to take their toll, however. Our new series reveal that published estimates, in some functions at least, are fundamentally incorrect.

The differences between spending series have real consequences. Different measures tell different stories, and we only hope that the new measures allow us to see things more clearly. Even as we have aimed to improve measures of functional spending in the UK, it is important to keep in mind that outlays are several times removed from budgetary policy itself. They are not spending 'policy' *per se*. What actually is spent ultimately reflects things that politicians cannot fully anticipate or manage. If we are interested in observing the effects of budgeting out in the world, the expenditures are of primary importance. If we are interested in capturing the Government commitment to programmes, budgetary *policy* is what matters. The problem is that finding reliable measures of actual policy in the UK is not easy; indeed, it may be impossible.

Such data are more readily available for the USA, where clear distinctions are made between budget authority, obligations and outlays. Appropriations bills in the USA specify an amount of budget authority that is available to an agency or department for obligation, i.e. for making commitments to spend money. When money is actually spent, an expenditure or outlay occurs, which can lag well behind appropriations decisions and obligations. Appropriations and outlays, thus, are two different things, and the former is the preferable measure of budgetary policy (see Wlezien and Soroka (2003)).

The reliance on outlays is an unfortunate fact of life for UK analyses, however. Our more accurate estimates of spending necessarily reflect a combination of recent budgetary decisions and the timing of actual expenditures, but they are still the best available measure of UK Governments' fiscal priorities. These series represent the first effort to develop temporally consistent measures of functional spending in the UK. Basic analyses indicate that they make a difference, especially in certain areas. The extent to which the data support or refute existing research on the causes and consequences of UK public expenditures remains to be seen.

Acknowledgements

We owe a special debt of gratitude to Michelle Meyer and Robert Bowles for their assistance in assembling our new budgetary data, to Russell Hubbard, Stuart Mitchell, Allan Ritchie, Eleanor Ball and Philippa Todd of Her Majesty's Treasury for their generous guidance throughout the process, and to many others for comments on the research, particularly Frank Baumgartner, Ian Budge, Erin Dyck, David Heald, Brian Hogwood, Peter John, Bryan Jones, Richard Parry, Stephen Smith, the Joint Editor and especially David Cox for suggesting the method of assessing the tails of the distribution that we use in Section 5. The data collection and analysis have been supported by a grant from the Nuffield Foundation.

References

Anscombe, F. J. and Glynn, W. J. (1983) Distribution of the kurtosis statistic b2 for normal samples. *Biometrika*, **70**, 227–234.

Baumgartner, F. and Jones, B. D. (1993) *Agendas and Instability in American Politics*. Chicago: University of Chicago Press.

Dawkins, R. (1986) *The Blind Watchmaker*. Harlow: Longman.

Heald, D. (1995) Steering public expenditure with defective maps. *Publ. Admin.*, **73**, 213–240.

Her Majesty's Treasury (2000) Resource budgeting and the 2000 Spending Review: explanatory note by HM Treasury. Her Majesty's Treasury, London. (Available from `http://www.hm-treasury.gov.uk./media/935/F6/83.pdf`.)

Her Majesty's Treasury (2005) Public expenditure statistical analyses 2005. *Cm 6521*. London: Her Majesty's Treasury.

Hogwood, B. W. (1992) *Trends in British Public Policy*. Buckingham: Open University Press.
John, P. and Margetts, H. (2003) Policy punctuations in the UK: fluctuations and equilibria in central government expenditure since 1951. *Publ. Admin.*, **81**, 411–432.
Jones, B. D. and Baumgartner, F. (1997) Does policy incrementalism stem from political consensus or from institutional gridlock? *Am. J. Polit. Sci.*, **41**, 1319–1339.
Jones, B. D., Baumgartner, F. and True, J. L. (1998) Policy punctuations: U.S. budget authority, 1947-1995. *J. Polit.*, **60**, 1–33.
Jones, B. D., Sulkin, T. and Larsen, H. A. (2003) Policy punctuations in American political institutions. *Am. Polit. Sci. Rev.*, **97**, 151–169.
Lindblom, C. (1975) *The Policy-making Process*. Englewood Cliffs: Prentice Hall.
McLean, I., Wlezien, C., Soroka, S. N., Bowles, R., Cameron, G., Coggins, R., Flamand, L., Hamill, H., Hepburn, C., Ingram, P., Jansky, R., Kemp, J., Levacic, R., Love, S., Noble, M., Owens, T., Ratcliffe, A., Spirling, A. and Wolf, A. (2003) Identifying the flow of domestic and European expenditure into the English regions. Office of the Deputy Prime Minister, London. (Available from `http://www.local.odpm.gov.uk/research/idenrprt.pdf`.)
Wildavsky, A. (1984) *The Politics of the Budgetary Process*, 4th edn. Boston: Little Brown.
Wlezien, C. (2004) Patterns of representation: dynamics of public preferences and policy. *J. Polit.*, **66**, 1–24.
Wlezien, C. and Soroka, S. N. (2003) Measures and models of budgetary policy. *Poly Stud. J.*, **31**, 273–286.