

Divisions of the United Kingdom House of Commons, from 1992 to 2003 and Beyond

David Firth and Arthur Spirling

Nuffield College, Oxford, UK

This version September 6, 2003

Summary. This article documents two new statistical resources for contemporary political study of the UK Parliament. The first is a set of freely available software tools, collectively known as *tapiR*, which allow House of Commons division lists to be extracted easily from the online pages of *Hansard* and organised into a convenient form for analysis. The second resource is voting data for all Members of Parliament, constructed by using *tapiR*, for two complete parliaments (1992–1997 and 1997–2001) plus the present one (2001 onwards).

Keywords: Legislative voting, Political science

1. Introduction

In political science the use of advanced statistical methods to analyse and represent patterns of legislative voting has become commonplace (see, for example, Ansolabehere *et al.*, 2001; Heckman and Snyder, 1997; Levitt, 1996; Poole and Rosenthal, 1991). However, research on the UK Parliament has been hampered by problems of data availability. Some historical studies have been made, in which investigators have gone to some lengths to construct datasets from the printed pages of *Hansard*, the official Parliamentary record; see, for example, the discussion of the Aydelotte data set in McLean and Bustani (2001) and McLean (2001). In other countries, notably the United States, voting records for the national legislature are routinely available in a convenient, spreadsheet-type form. In contrast, the pages of *Hansard* provide only individual division lists in textual form; even the online *Hansard* pages, which are available for all House of Commons debates since 1995, cannot be used directly for analysis, as they contain inconsistencies in the names used for Members of Parliament (MPs), extraneous markup codes, etc. For the statistical study of MPs' participation in Early Day Motions (e.g., Nason, 2001) a convenient database is available (at <http://edm.ais.co.uk/>); the aim of the work reported here is to provide a similarly useable resource for the systematic study of voting in House of Commons debates.

The resource is in two parts: data on House of Commons voting for the period 1992 to 2003; and software tools to enable future House of Commons divisions, from May 2003 onwards, to be added easily by accessing the relevant pages of *Hansard* online and incorporating the votes of each MP into a single rectangular dataset.

The software tools take the form of a standard package, for the *R* statistical computing environment (Ihaka and Gentleman, 1996), which allows for convenient distribution and installation on a variety of platforms. The package is named *tapiR*, standing for *tools for accessing parliamentary information in R*. The *tapiR* software has full online documentation

Address for correspondence: David Firth, Nuffield College, Oxford OX1 1NF, UK.
E-mail: david.firth@nuffield.ox.ac.uk

included, in the customary style for published *R* packages. The description given in this article provides a more gentle and discursive overview than the necessarily rather formal documentation that comes with the *tapiR* package itself.

The voting datasets are available both as part of the *tapiR* package, for use with *R*, and separately as spreadsheet-type ‘comma-separated values’ (.csv) files for reading into other common data-handling systems such as *Stata*, *SPSS* and *Excel*.

In the *Sixth Report* (1998) of the House of Commons Select Committee on Modernisation, which may be found at <http://www.parliament.the-stationery-office.co.uk/pa/cm199798/cmselect/cmmodern/779/77902.htm>, the possibility of an electronic voting system is discussed, which could go a long way towards eliminating the need for a project such as ours. However, as the *Sixth Report* makes clear, such an electronic system will not be introduced in the foreseeable future.

2. Votes for the period 1992 to 2003

Three separate datasets are provided for the 1992–97 parliament, the 1997–2001 parliament, and the current parliament which began in 2001. The data are in spreadsheet format, with rows representing MPs and columns representing House of Commons divisions. The first column after the MP names identifies each MP’s political party; in the few cases where an MP’s affiliation changes during the period, party membership at the end of the parliament is recorded.

	party	div001.010626	div002.010627	div003.010627
Abbott, Diane	Lab	n	n	n
Adams, Irene (Pai)	Lab	n	n	n
Ainger, Nick	Lab	n	n	n
Ainsworth, Bob (Cov)	Lab	n	n	n
Ainsworth, Peter (E S)	Con	–	y	–
Alexander, Douglas	Lab	–	n	n
Allan, Richard	LD	y	n	y
Allen, Graham	Lab	n	n	n
Amess, David	Con	y	y	–
Ancram, Michael	Con	y	y	–

Table 1: A small part of the voting data from June 2001

Table 1 illustrates the format of the data. Shown in Table 1 are votes for the first ten MPs alphabetically, in the first three divisions of the session which began after the June 2001 General Election. Divisions are identified by the Division Number as reported in *Hansard*, and the date on which the division took place. Votes are either ‘y’ or ‘n’, signifying respectively Ayes and Noes as recorded in the *Hansard* division lists (including the Tellers for each side). A dash means that the MP did not vote, i.e., was in neither the Ayes nor the Noes. Note that it is not possible to distinguish between different types of non-voting, the three most important such being absence from the House, deliberate abstention when present, and ‘pairing’ in which MPs on opposite sides of an issue agree that neither need vote as their votes would cancel. Occasionally *Hansard* records an MP in both the Ayes and Noes lists for a division, and in such instances we have used the code ‘b’ (for ‘both’); this can happen either because the MP went through both the Aye and the No lobbies, for

example because of a change of mind or the realisation of a mistake on their part, or as the result of an error in the published division lists.

Each of these three datasets is made available as a standard ‘comma-separated values’ (.csv) file of the kind that can be read by many standard computer programs, including *Stata*, *SPSS* and *Excel*. In *Stata* the data for 1997 to 2001, for example, can be read in by using the command

```
insheet using Votes9701.csv
```

and in *R* the same data can be read into a data-frame object by

```
Votes9701 <- read.csv("Votes9701.csv", row.names = 1)
```

where `row.names=1` means that the first column of the file, the MP names, will be used as row names in the resultant data frame.

The files are rather large: in 1992–97 there were 668 distinct MPs and 1285 divisions, and in 1997–2001 there were 671 MPs and 1279 divisions. Our experience with *Stata* and *R* has been that this is unproblematic; but with other software packages, especially those which impose limits on the number of data columns, some external pre-processing may be necessary, for example to select a subset of the divisions for analysis.

3. The *tapiR* package

3.1. The main functions of *tapiR*

All divisions up to 8 May 2003 are already recorded in the files described above. This section outlines a set of software tools which allow divisions taking place after that date to be added easily. The *tapiR* package is available on the *Comprehensive R Archive Network* (CRAN) for straightforward installation into the user’s local implementation of *R*. The *tapiR* package contains *R* functions to perform data-handling tasks described below, as well as containing data on divisions in the period June 2001 to May 2003.

There are three main tasks involved in converting one or more *Hansard* division lists into columns of a spreadsheet as illustrated in Table 1:

- (i) Obtain a complete ‘master’ list of unambiguous MP names for the parliamentary session. Some MPs share the same name, and our approach to eliminating ambiguity in such cases is to append the first three characters of an MP’s constituency name; thus Peter Ainsworth’s constituency, which *Hansard* normally records as ‘(E Surrey)’, becomes ‘(E S)’.
- (ii) Download the relevant page(s) from the online *Hansard*, extract the division-list information and strip out redundant information such as hypertext markup (HTML) codes.
- (iii) For each MP name in each division list obtained, identify the MP in the ‘master’ list and record their vote in the spreadsheet, or ‘-’ if no vote, as shown in Table 1.

For the current parliament beginning in June 2001, task (i) has been completed already: the MP names collected from division lists between 2001 and 2003 are contained in a dataset named `MPnames0103` in *tapiR*. This was achieved using the *tapiR* function `collect.MPnames`, which initially produces an exhaustive list with many ambiguities; for example, *Hansard* sometimes makes spelling errors in MPs’ names, resulting in the same real name appearing more than once but with different spellings. All such ambiguities

4 D. Firth and A. Spirling

were removed, by a combination of corrections (of minor errors in the division lists) and rule-making (such as the rule that ‘Ainger, Nicholas’ refers to ‘Ainger, Nick’ in the master list). Alteration of `MPnames0103` in the future should be needed only in the event of the addition of an MP at a by-election during the term of the parliament, in which case an extra row should simply be added to the data. The `MPnames0103` dataset also stores the three-character abbreviation of the constituency name where used, and the party affiliation of each MP.

For task (ii), *tapiR* provides the functions `find.divisions` and `get.divisions`. The first of these searches *Hansard* for divisions which took place on a specified date or dates in a specified parliamentary session, and returns the URL for each division list found; while `get.divisions` visits a given set of URLs and extracts the full division lists. To illustrate, consider the capture of data for Division Nos. 75–79 on 4 February 2003, which were all ‘free’ votes on House of Lords reform options; these votes are analysed in McLean *et al.* (2003). These divisions are captured simply by

```
> URLs4Feb03 <- find.divisions(0203, 030204)
> Divs4Feb03 <- get.divisions(URLs4Feb03)
```

The resultant object `Divs4Feb03` is a five-component list, with each component corresponding to one division. The five components are named ‘div075.030204’ to ‘div079.030204’, and each of these five is itself a list with three components named ‘url’, ‘ayes’ and ‘noes’, the last two being the division lists. Thus, for example,

```
> Divs4Feb03$div075.030204$ayes
 [1] "Anderson, Janet (Rossendale & Darwen)"
 [2] "Banks, Tony"
 [3] "Barnes, Harry"
 ...
[173] "Tellers for the Ayes:"
[174] "Jim Fitzpatrick and"
[175] "Dan Norris"
```

Finally, for task (iii) above, the division lists are checked against the master list of MP names, and votes recorded, using the *tapiR* function `make.votesheet`:

```
> Votes4Feb03 <- make.votesheet(Divs4Feb03, MPnames0103)
> Votes4Feb03
      party div075.030204 div076.030204 div077.030204 ...
Abbott, Diane      Lab      -            n            y
Adams, Irene (Pai) Lab      -            y            n
Ainger, Nick       Lab      n            y            n
...
```

The three functions `find.divisions`, `get.divisions` and `make.votesheet` together provide most of the functionality of *tapiR* for routine use. Sometimes `make.votesheet` may complain that a name found in a division list is not present in the master list, in which case either the division list should be corrected, or a new ‘rule’ should be added (if, for example, an MP has come to be known frequently by some new style, such as ‘Bob’ rather than ‘Robert’). The *tapiR* auxiliary function `namerules.9203` encodes all such rules as used for divisions which took place between April 1992 and May 2003.

Once a ‘votesheet’ object has been created for a set of divisions, it can be operated upon in the standard ways for data frames in *R*. In particular, the combination of votes for two or more sets of divisions is easily achieved using *R*’s `cbind` function. To make the voting data available for analysis in other systems, a standard ‘comma-separated values’ (`.csv`) file can be written by using the *tapiR* function `write.csv`; for example,

```
> write.csv(Votes4Feb03, file = "Votes4Feb03.csv")
```

One further *tapiR* function worth noting here is `write.votesheet`, which makes a compact representation of the voting data either for on-screen display or in a file. For example,

```
> write.votesheet(Votes4Feb03)
                party votes
Abbott, Diane      Lab -nyyy
Adams, Irene (Pai) Lab -ynnn
Ainger, Nick       Lab nynnn
Ainsworth, Bob (Cov) Lab nynnn
Ainsworth, Peter (E S) Con -----
Alexander, Douglas Lab nnyyy
...
```

The codes ‘y’, ‘n’, etc., can be changed as desired. For example, certain analytic computer programs commonly used in political science take as their input (1,6,0) in place of (y,n,-); to write such a file, with code ‘b’ also replaced by ‘-’ wherever it occurs, the usage of `write.votesheet` would be of the form

```
> write.votesheet(Votes4Feb03, file = "Votes4Feb03.txt", keep.b = FALSE,
                 aye = "1", no = "6", novote = "0")
```

3.2. Why *R*?

The decision to implement *tapiR* in the *R* statistical computing environment was influenced by a number of factors, including:

- (a) *R* is freely available for almost all modern computing platforms;
- (b) *R* provides full programming facilities, flexible data-handling functions and direct access to web-based data sources;
- (c) the number of users is large and increasing rapidly, including in political science where *R* has been adopted for the training of graduate researchers by both the ICPSR Michigan Summer School and the ESRC Oxford Spring School;
- (d) the availability of standardized systems for online documentation of a package such as *tapiR*, and for distribution via the Comprehensive R Archive Network (CRAN, at <http://cran.r-project.org>).

The application of *tapiR* as described in Section 3.1 does not require extensive knowledge of *R*, but users of *tapiR* will need an understanding of the basics, to the level of the document *An Introduction to R* which should be present in every installation of *R*.

4. Further work

4.1. Maintenance

The *tapiR* software will be maintained as a standard *R* package on CRAN, and updated as necessary, for example when new MPs enter the House or when a sitting MP becomes known by a new name. The voting data for the whole of the current term, beginning in June 2001, will be made available in the same format as the data for 1992–97 and 1997–2001, after the next General Election. This commitment assumes that *Hansard* continues to publish division lists online in a format similar to that which has been used since 1995. It is likely, given the availability of *tapiR* and its ease of use, that the authors will continue to make such data available also for future parliaments, while ever there is a need.

The datasets themselves, in the common *.csv* format, will be available from the RSS datasets website, currently at <http://www.blackwellpublishing.com/rss/>. [Assuming that this article is accepted by the Journal!] At the time of writing they are available from <http://www.stats.ox.ac.uk/~firth/tapiR/>.

4.2. Earlier parliaments

The *Hansard* division lists prior to 1988 have not yet been published in electronic form; for the period 1988 to 1995 they are available on CD-ROM. For divisions prior to 1988, then, the application of *tapiR* to produce a useable rectangular dataset would first require the optical scanning of printed pages from *Hansard*, and then checking and cleaning of the resulting files. In contrast to the direct use of *tapiR* to ‘scan’ online pages of *Hansard* as described in Section 2.1 above, this would demand a great deal of detailed attention from one or more skilled people. If suitable funding can be obtained this will be pursued further.

4.3. Additional attributes of divisions and MPs

The voting datasets introduced here, although providing a rich new resource for contemporary empirical research on the House of Commons, could be made still more valuable by the addition of:

- (a) further information on MPs, such as gender, age, time since first entering the House, whether a (shadow) Minister, etc.; and
- (b) further information on divisions, such as the policy area of the debate (economy, defence, etc.), whether the division was a ‘free’ (i.e., not whipped) vote, whether the division was connected with a Government-sponsored motion, etc.

The systematic addition of such contextual data would require substantial further work by a political researcher familiar with the period in question.

4.4. Scaling methods

This new data resource will enable more sophisticated analyses of British parliamentary behaviour than have hitherto been possible. In particular, it allows the application of scaling techniques of the kind commonly used in political science for this type of data (e.g. Poole, 2003; Poole and Rosenthal, 1991, 1997). The most frequently used software packages which implement such methods, NOMINATE (Poole and Rosenthal, 1987) and OC (Poole, 2000) are widely used and discussed in the US academic community. Unfortunately, preliminary

investigation indicates that their utility with recent British data may be rather limited. There are two main reasons:

- (a) In the UK parliament most MPs vote with the rest of their party, usually under instruction from party whips, on most issues. Thus most parliamentary voting is informative only about party membership, which is known *a priori*, rather than about ideology or other MP attributes of interest (Londregan, 2000).
- (b) The standard scaling approaches assume monotonicity: along any notional ideological dimension, ideal voting will be in one direction ('aye', say) for MPs to the left of some point of the scale, and in the other direction ('no') for those to the right of that point. This is strongly at odds with the pattern observed on many occasions in the UK parliament, where a motion may be supported by those in the centre but opposed, most likely for different reasons, by members to both the left and the right (McLean and Spirling, 2003).

Point (a) suggests that most attention should be paid to 'free' votes, in which party whips make no instruction, or to divisions in which a substantial fraction of MPs rebel against the party whip. Point (b) demands the development of novel approaches to scaling which accommodate voting patterns of the type described; this more methodological project is under way, but is outside the scope of the present paper.

Acknowledgements

The authors thank Iain McLean for helpful conversations, and Keith Poole for advice on the use of his OC (optimal classification) software. The work of A.S. was supported by the Economic and Social Research Council.

References

- Ansolabehere, S., Snyder, J. and Stewart, C. (2001) Candidate positioning in U.S. House elections. *American Journal of Political Science*, **45**, 136–159.
- Heckman, J. and Snyder, J. (1997) Linear probability models of the demand for attributes with an empirical application to estimating the preferences of legislators. *Rand Journal of Economics*, **28**, S142–S189.
- Ihaka, R. and Gentleman, R. (1996) R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, **5**, 299–314.
- Levitt, S. (1996) How do senators vote? Disentangling the role of voter preferences, party affiliation, and senator ideology. *American Economic Review*, **86**, 425–441.
- Londregan, J. (2000) Estimating legislators' preferred points. *Political Analysis*, **8**, 35–56.
- McLean, I. (2001) *Rational Choice and British Politics: An Analysis of Rhetoric and Manipulation from Peel to Blair*. Oxford: Oxford University Press.
- McLean, I. and Bustani, C. (2001) On providence, potatoes, and Peel. *Political Studies*, **49**, 110–111.

- McLean, I. and Spirling, A. (2003) UK OC OK? A note on interpreting optimal classification scoring for the United Kingdom. *Tech. rep.*, Nuffield College, Oxford. URL <http://www.nuff.ox.ac.uk/Politics/papers/2003/OC0McLean.pdf>.
- McLean, I., Spirling, A. and Russel, M. (2003) None of the above: The UK House of Commons votes on reforming the House of Lords, 2003. *The Political Quarterly*, **74**, 298–310.
- Nason, G. (2001) Early Day Motions: Exploring backbench opinion during 1997–2000. *Tech. rep.*, Department of Mathematics, University of Bristol. URL <http://www.stats.bris.ac.uk/pub/ResRept/2001/11.pdf>.
- Poole, K. (2000) Non-parametric unfolding of binary choice data. *Political Analysis*, **8**, 211–237.
- (2003) Changing minds? Not in Congress! *Tech. rep.*, Department of Political Science, University of Houston. URL <http://voteview.uh.edu/chminds/chminds2.htm>.
- Poole, K. and Rosenthal, H. (1987) Analysis of Congressional coalition patterns: A unidimensional spatial model. *Legislative Studies Quarterly*, **12**, 55–75.
- (1991) Patterns of Congressional voting. *American Political Science Review*, **35**, 228–278.
- (1997) *Congress. A Political-Economic History of Roll Call Voting*. New York: Oxford University Press.