Modelling Political Popularity:
An Analysis of Long Range Dependence in Opinion Poll Series

by

David Byers

James Davidson

(Cardiff Business School)

and

David Peel

(University of Liverpool)

Revised, December 1996

Summary

A simple model of political popularity, as recorded by opinion polls of voting intentions, is proposed. We show that, as a consequence of aggregating heterogeneous poll responses under certain assumptions about the evolution of individual opinion, the time series of poll data should exhibit long memory characteristics. In an analysis of the monthly Gallup data on party support in the UK, we confirm that the series are long memory, and further show them to be virtually pure 'fractional noise' processes. An explanation of the latter result is offered. We study the role of economic indicators in predicting swings in support, perform event analyses, and use our estimates to generate post-sample forecasts to April 1997.

Keywords: Political popularity, opinion polls, long memory, fractionally integrated processes

Address for correspondence:

James Davidson, Cardiff Business School, Colum Drive, Cardiff CF1 3EU

email: DavidsonJE@cardiff.ac.uk

1. Introduction

Following the seminal work of Goodhart and Bhansali (1970), numerous studies have examined the evolution of voting intentions, as measured by opinion polls, and in particular the relationship between political popularity and economic variables such as inflation and unemployment. See for example Pissarides (1980), Minford and Peel (1982), Schneider and Frey (1988), Rogoff and Sibert (1989), and Gärtner (1994). An empirical issue of particular relevance to the present study is the degree of persistence in political popularity. Building on the rational expectations version of the permanent income hypothesis due to Hall (1978), Holden and Peel (1985) (see also Chrystal and Peel 1987) argued that the effect of news about the economy on voting intentions would be permanent. The practical implication of their model is that the time series of opinion data should behave like a random walk, with the autoregressive-moving average (ARMA) representation of the time series containing an autoregressive root of unity. Such processes are nonstationary, and exhibit no mean-reversion tendencies.

However, further analysis of the UK data in Byers (1991) rejected the unit root hypothesis in favour of stationary ARMA models, although with autoregressive coefficients close to unity. Similar results are reported by Scott, Smith and Jones (1977). Such models would imply that the effect of news on voting intentions, although it could be quite persistent in practice, is in principle transitory. These findings are reflected in Figs. 1(a) and 2(a) below, which plot the monthly Gallup poll data for Conservative and Labour support for the period September 1960 to May 1995 (417 observations). The data have been logistically transformed so that they are distributed on the whole real line, with the value zero corresponding to 50% support. A random walk representation is admissible in this case, and the series do indeed exhibit highly persistent swings in support, but they also appear to be eventually mean reverting.

In this paper we present a new model of popularity that reconciles theory and observation, deriving its special features from the effect of aggregating heterogeneous poll responses. Suppose for simplicity that voters fall into two stylised categories, the committed and the uncommitted. The committed individuals are those with strong party allegiances. They are motivated by principle, or ideological conviction, and their voting intentions are generally insensitive to news. They will support their party of choice through good times and bad. The uncommitted individuals, on the other hand, who are usually called ‘floating voters’, tend to award their votes on the basis of performance. Newspaper headlines can sway their voting intentions considerably, although by the same token, the effect of news is typically transitory, and will tend to average out in the long run. The current voting intentions of the floating voters are on the whole a poorer predictor of future voting intentions than those of the committed voters.

These features of the voting population are captured in our model by assuming that the logarithm of the odds in favour of an individual expressing an intention to vote for a particular party evolves as a first-order autoregressive (AR) process, with a parameter that takes a value close to unity for committed voters, and substantially below unity for the floating voters. Consider a population of voters, each characterised by a particular value of the AR coefficient on the interval [0,1]. If the distribution of these coefficients is as we hypothesise, a result due to Granger (1980) implies that the average of a large sample of individual voting intentions (after a logistic transformation) should behave similarly to a so-called fractionally integrated process. This means that it should exhibit long memory, unlike a stationary ARMA process, in spite of being eventually mean-reverting. (See Beran 1992 for a recent survey of long memory processes, and also Haslett and Raftery 1989 for an application of this model to a different problem.) Statistical analysis of the Gallup series lends support to our model. Such a model would imply that changes in party fortunes, such as that experienced by the Labour Party in the 1980s and more recently by the Conservative Party, are difficult although not impossible to reverse once established. An important technical implication of the model is that regressions that appear to explain opinion poll results in terms of economic variables could be spurious.

The paper is structured as follows. In Section 2 we set out formally the model of voting intentions, and derive from this the distribution of opinion poll data. Section 3 discusses the specification and estimation of the election cycle, an important stylised component of the poll series. In Section 4 we report empirical tests and estimates of the fractional integration model. We discuss and interpret our key empirical finding, that the time series appear to follow a virtually pure fractional noise process. Section 5 analyses the fractionally differenced series, reporting regressions on economic indicators and event analyses, and Section 6 concludes the paper by reporting some forecasting exercises based on the model.

Fig. 1. Conservative support, Sept. 1960-May 1995
(Logistic transform of the Gallup poll data)

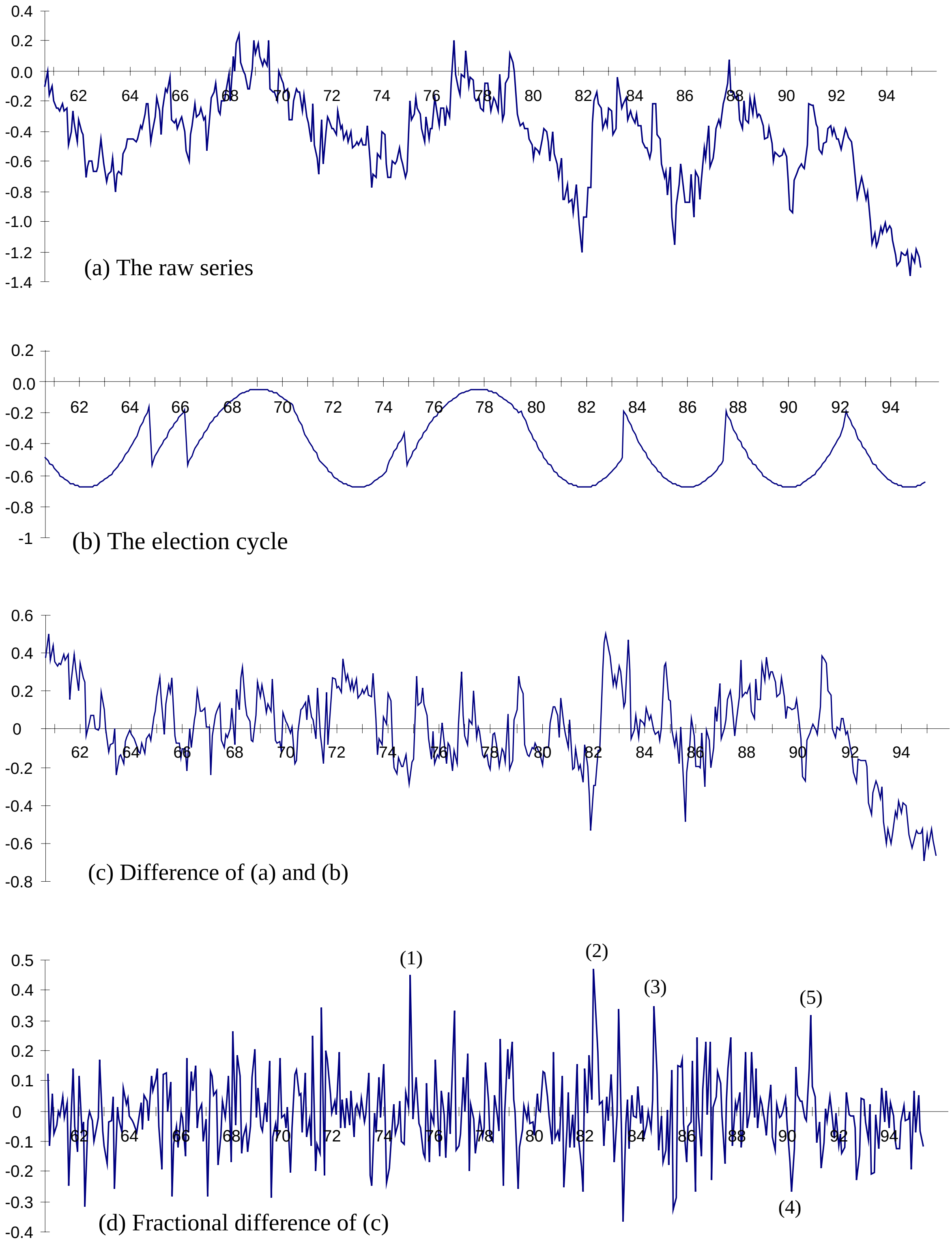
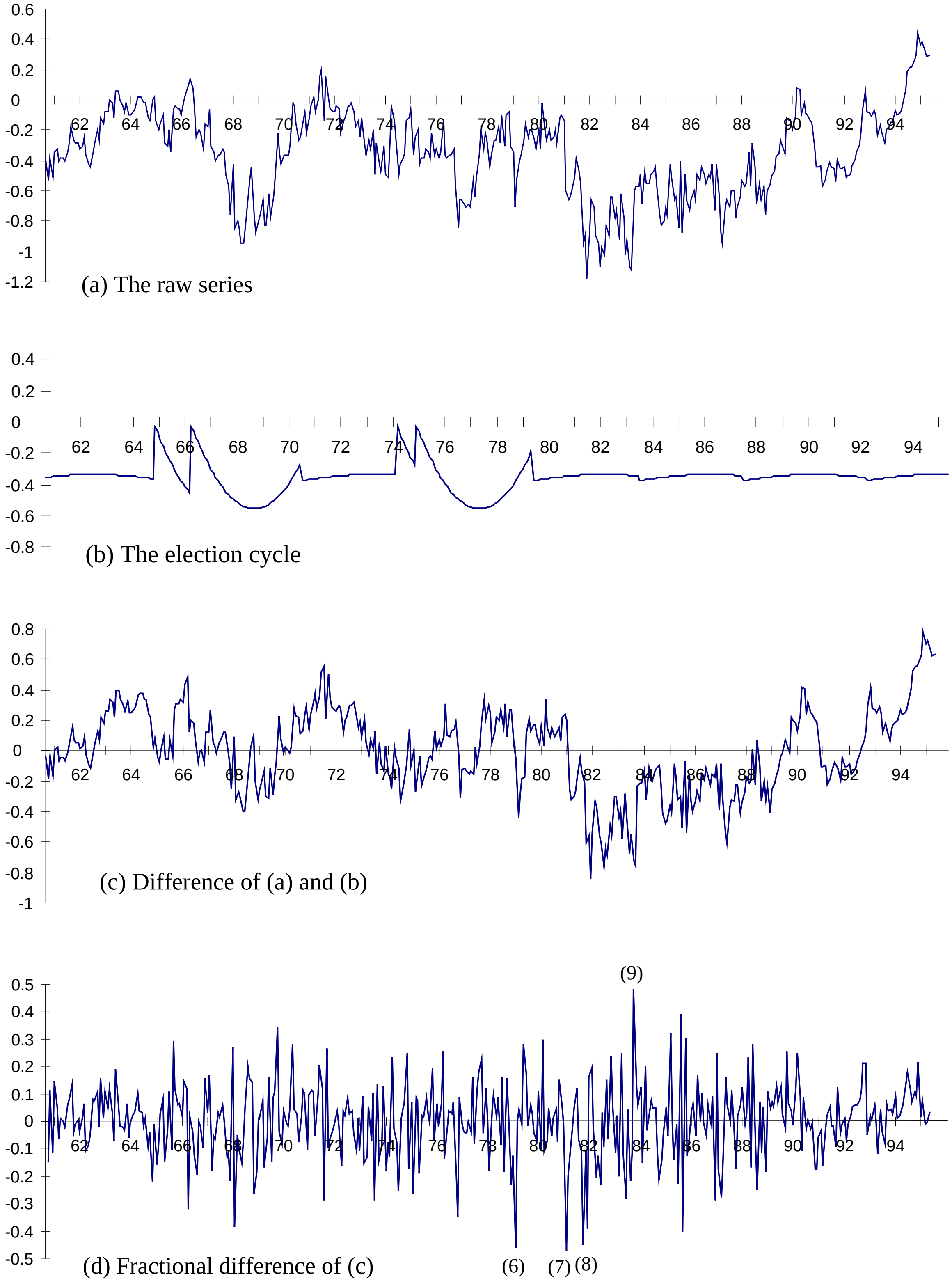


Fig. 2. Labour support Sept 1960-May 1995
(Logistic transform of the Gallup poll data)



2. A Model of Persistence in Poll Data

Let p_t^i denote the probability that voter *i* supports (say) the Conservative Party, in a survey of voting intentions conducted on date *t*. Think of this as a latent variable representing strength of support for the Conservatives. We think of the voter's choice as a binary one, in this sense, although of course, any voter who decides not to vote Conservative then has a further choice between the other parties on offer. The same analysis can be performed in respect of Labour support, or any other party for that matter.

While p_t^i is a latent variable, it is the expectation of an observable one, the actual binary response, X_t^i, of the voter to a poll of voting intentions, equal to one with probability p_t^i and zero with probability $1 - p_t^i$. The law of large numbers implies that the sample mean response in a sample of *N* voters, $\overline{X}_t = N^{-1}\sum_{i=1}^{N} X_t^i$, converges in probability to the same limit as the average probability, $\overline{p}_t = N^{-1}\sum_{i=1}^{N} p_t^i$. (See, for example, Davidson 1994, Thm 19.1). Similarly, consider the log-odds in favour of a Conservative vote,

$$Y_t^i = \log \frac{p_t^i}{1 - p_t^i}. \tag{2.1}$$

Appealing to Slutsky's Theorem on the convergence in probability of a continuous function of a convergent stochastic sequence (see, e.g. Davidson *op. cit.* Thm 18.8) $\overline{Y}_t = N^{-1}\sum_{i=1}^{N} Y_t^i$ will be arbitrarily close in probability, when *N* is large, to $\log\{\overline{X}_t/(1-\overline{X}_t)\}$. The latter (observed) variable may be expected to follow the same stochastic process as $\overline{Y}_t$ to the same degree of approximation.

We hypothesise that

$$Y_t^i = y_t^i + C_t^i \tag{2.2}$$

where y_t^i evolves according to

$$y_t^i = \alpha^i y_{t-1}^i + \varepsilon_t^i \tag{2.3}$$

where $0 \le \alpha^i \le 1$ and ε_t^i is a shock variable. We assume that this represents the impact of news, and we discuss its structure in greater detail below. C_t^i is a term representing the effect of the 'election cycle', whose specification we discuss in Section 3.

For the voters we have described as committed, we assume that $\alpha^i \approx 1$. For these individuals y_t^i is close to a random walk, and as such has no tendency to revert to zero, even if $E(\varepsilon_t^i) = 0$. It may increase or decrease without limit, corresponding to p_t^i approaching 1 or 0. The logistic form of the response function implies that such voters, given an initial high probability of supporting a particular party, are unlikely to change their support in the future. Floating voters, by contrast, are those with α^i significantly smaller than unity. They change their allegiance freely, in the face of news. Their behaviour is relatively volatile because, given the logistic form of the response, news has a greater impact on p_t^i when this is close to 1/2.

Our model is very simple, but its description of voters' behaviour has some interesting features. The committed voters may not immediately react to news, but they do not ignore it. While taking care not to push the parallel too far, we note that they have something in common with the rational consumers postulated in the Hall (1978) model of consumption, whose permanent income moves as a random walk. The link between rationality and our model of commitment is that in both cases, information is incorporated permanently into future behaviour, and not discounted with the passage of time. Committed voters weigh the historical context equally with current performance, in contrast to floating voters for whom current performance is the sole criterion. Since even a random walk may occasionally change sign, the committed voter is not excluded from undergoing a Damascene conversion, following a persistent flow of unfavourable news. However, if we compare the two cases by considering what would happen in a steady state, in the absence of news, the committed voter persists with his established preference, whereas the floating voter reverts to voting at random, with $p_t^i \to 1/2$.

We have spoken of two groups for clarity of exposition, but we do not mean to imply that the distribution of the α^is in the voting population is bimodal. In fact, we assume that they have a continuous distribution on [0,1] in which part of the mass is concentrated near 1. A class of distributions that can reproduce the required characteristics is the modified Beta(*u*,*v*), given by

$$dF(\alpha) = \frac{2}{B(u,v)}\alpha^{2u-1}(1-\alpha^2)^{v-1}d\alpha\,, \tag{2.4}$$

with $0 < v < 1$. A particular case of this density is sketched in Fig. 3. In this class of distributions zero mass is assigned to the point 1 itself, but unit and near-unit roots give rise to similar behaviour except in the very long run.

Next consider the specification of the shock variable. We assume that this has the form

$$\varepsilon_t^i = \sum_{j=1}^{J}\beta_j^i W_{jt} \tag{2.5}$$

where the real random variables W_{jt}, $j = 1,..., J$, (where J is an unspecified large, but finite, number) are cardinal representations of the various categories of ‘news’ of concern to voters. These might include unexpected changes in economic indicators such as growth and inflation, as well as measures of the impact of national political events. The variables α^i and $\beta_1^i,\ldots,\beta_J^i$ are assumed to have a joint distribution over the voting population that is fixed with respect to time, and in which α^i is distributed independently of the other components. Equation (2.5) embodies our assumption that voters all react to the same news, but it allows them to perceive and respond to it in different ways. A rise in interest rates, for example, is good news to lenders but bad news to borrowers. The distributions of the β_j^i must generally be supported on both positive and negative values.

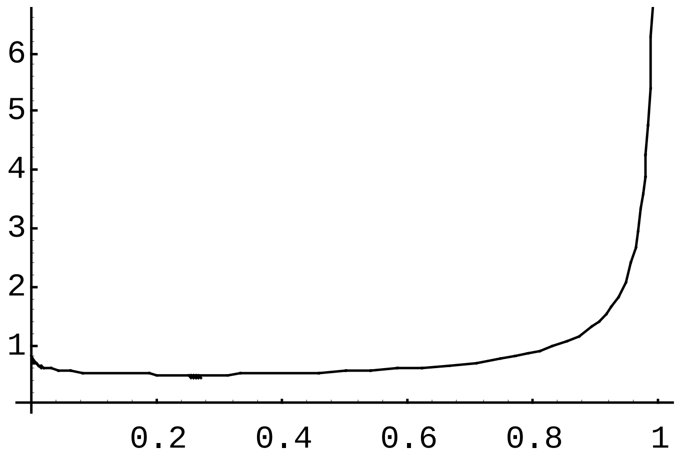


Fig. 3. The Beta(0.45, 0.3) density function

Under these assumptions, we can apply the results of Granger (1980) on the aggregation of heterogeneous AR(1) processes. Letting L denote the lag operator, such that $L^k X_t = X_{t-k}$, we will use the general notation $a(L) = \sum_{k=0}^{\infty} a_k L^k$ to denote a lag polynomial such that $a(L)X_t$ is an infinite-order moving average of the sequence X_t. Defining $\bar{y}_t = N^{-1}\sum_{i=1}^{N} y_t^i$, Granger’s results imply that to a good approximation when N is large,

$$\bar{y}_t \approx \sum_{j=1}^{J} E(\beta_j)\alpha(L)W_{jt} \tag{2.6}$$

where the $E(\beta_j)$ are constants, and

$$\alpha(L) = \int \frac{1}{1-\alpha L}\, dF(\alpha). \tag{2.7}$$

Granger points out (1980, p. 235) that when the density takes the form (2.4) with $0 < v < 1$, the coefficients of the lag polynomial $\alpha(L)$ have the form

$$\alpha_k = \frac{B(u+k/2,v)}{B(u,v)} = O(k^{-v}). \quad (2.8)$$

It is well-known (see for example Granger and Joyeux 1980, Hosking 1981) that this rate of decline of the moving average coefficients is shared with 'fractionally integrated' processes, which have the representation

$$x_t = (1-L)^{-d} u_t \quad (2.9)$$

where u_t is a stationary process, for a non-integer value of d in the range $0 < d < 1$. For example, if u_t is a white noise process, x_t in (2.9) is called a pure fractional noise, and has an infinite-order moving average representation in terms of the standard binomial expansion of $b(z) = (1-z)^{-d}$. The coefficients are shown by Granger and Joyeux (1980) to have the form

$$b_k = \frac{\Gamma(k+d)}{\Gamma(d)\Gamma(k+1)} = O(k^{d-1}). \quad (2.10)$$

Thus, the length of memory of the process, as measured by the weights assigned to remote time periods in the average, varies directly with d. The formula in (2.10) is also valid for $-1 < d < 0$, which can be thought of as the case obtained by the simple differencing of a fractionally integrated process.

If u_t has a stationary invertible ARMA(p,q) representation, x_t is sometimes called a ARFIMA(p,d,q) process (see Diebold and Rudebusch 1989). Extended to cover the more familiar ARIMA(p,d,q) class by including the cases $d = 0$ and $d = 1$, processes in this class are referred to generally as I(d). When $d < 0.5$ they are stationary, with finite variance and autocorrelations that decline like k^{2d-1}, whereas when $d \geq 0.5$ they are nonstationary. The hyperbolic decay of the memory of a fractionally integrated process contrasts with the eventually exponential decay of the stationary ARMA, or I(0), process, and enables ARFIMA processes to model dependence between observations at long range. But at the same time, in contrast to the I(1) (or unit root) case, these processes exhibit eventual mean reversions.

Comparing (2.8) with $v < 1$ and (2.10), assuming that W_{jt} ~ I(0) for each j, it appears that $\alpha(L)$ might be adequately modelled by an integrating filter of order $d = 1 - v$, so that $\bar{y}_t$ ~ I(d). Note that the condition on v is just what is needed to produce a distribution in the class (2.4) with mass concentrated close to 1. The value of u is not critical for this result, but determines the shape of the distribution in the region close to 0, and hence the characteristics of the floating-voter population. Indeed, the assumption of a Beta distribution does not appear to be critical either, since a wide range of shapes can be generated for different choices of u and v. However, it should be noted that assigning a positive probability mass to the point 1 in the aggregate would imply that $\bar{y}_t$ ~ I(1).

These results are derived indirectly and cannot tell us whether $\bar{y}_t$ has a low-order ARFIMA representation, or what the parameters of that process might be. There is an identification problem here, in that the structure of the aggregate process depends on both u and v, and on the structure of the W_{jt} processes. Since we interpret the latter as 'news' variables, we have a particular interest in the case where the W_{jt} processes are white noise, but without knowing u, we cannot infer this fact from a study of the aggregate process. However, as we will show below, the white noise assumption has empirical plausibility.

This is our basic model. The assumptions are all necessary components of the model of aggregation over agents, but in evaluating them, take care to note that the model describes only the evolution of the latent variable p_t^i, measuring attitude to a national party. Equation (2.5) simply says that individual or local factors (such as would need to bear both the i superscript and the t subscript) are not incorporated into the voter's stance on national politics. Thus, loyalty to a popular MP might cause voters to override their view of the party nationally, but his successor would not expect to inherit his personal vote. Such factors may of course influence the actual poll responses X_t^i, but these variations are assumed to average out in the aggregate poll result. The available evidence suggests that in an application to the Gallup data, N is large enough in practice to permit sampling variation to be neglected. Gallup's sample sizes are of the order of 1000, and although the polls are taken by the method of quota sampling, the evidence in Smith (1996) suggests that the effect of sample design on sampling variance is not large. We also note that quota sampling introduces a possible time dependence in the sampling variation (see Scott, Smith and Jones 1977) but for the same reasons, would not expect this significantly to affect our result.

3. The Election Cycle

The basic model needs modification to take account of the tendency of a party's support to depend on the proximity of an election, which has been found important in previous studies; see for example Goodhart and Bhansali (1970), Miller and Mackie (1973), and Frey and Schneider (1978). The election cycle can be a dominant feature of the data, see Fig. 1(a), in particular. This is the purpose of the term C_t^i in equation (2.2).

Specifically, we distinguish ‘in power’, ‘honeymoon’ and ‘anticipation’ effects. The relationship between news and shifts in popularity is bound to depend on whether the party is in power or in opposition. Moreover, it is well known that support for the ruling party tends to be high in the months following an election, falls to a mid-term trough after 18 month to 2 years, but then picks up again as the expiry date of the current parliament approaches and the likelihood of an election being called rises. (Recall that elections in Britain inside the maximum five-year term are held at the discretion of the party in power.) The amount of the current term expended appears to exert at first a negative and later a positive effect on the voter’s preferences for the ruling party. It is easy to rationalise these swings, which are also seen in by-election results, in terms of voters who normally support the governing party registering a ‘safe’ protest at unpopular policies. The main opposition party experiences the obverse effect, although to a lesser degree, for it is a familiar fact of British electoral life that the small parties (Liberals and nationalists) pick up much of the mid-term protest vote.

Our approach to modelling the cycle is to construct dummy variables, synchronised to the relevant dates. Let D_t = 1 when the party in question (say, the Conservatives) is in power, and 0 otherwise. Defining T_t as the elapsed time since the last election, we model the effect of the electoral cycle on voter *i* as

$$\begin{aligned} C_t^i = \gamma_1^i D_t + \gamma_2^i (1 - D_t) + \gamma_3^i D_t T_t + \gamma_4^i (1 - D_t) T_t \\ + \gamma_5^i D_t T_t^2 + \gamma_6^i (1 - D_t) T_t^2 \end{aligned} \tag{3.1}$$

The coefficients γ_1^i and γ_2^i (the intercepts) can be interpreted as representing the unconditional probability of support in the month of an election, depending on whether the party has won the election and is in power (D_t = 1) or lost it and is in opposition (1 – D_t = 1). We would expect that $\gamma_1^i > \gamma_2^i$, which is the baseline ‘in power’ effect. The model allows this mean support to change as the term of government proceeds, at rates measured by γ_3^i for the party in power and by γ_4^i for the opposition. We would expect that $\gamma_3^i < 0$ and $\gamma_4^i > 0$ (the ‘honeymoon’ effect), but the trend is eventually reversed as the end of the term approaches and an election becomes more likely (the ‘anticipation’ effect), captured by the terms in the square of the trend, with anticipated signs $\gamma_5^i > 0$ and $\gamma_6^i < 0$. Since the actions of the party in power are more influential, for good or ill, than the oppositions’, we would also expect that $|\gamma_3^i|$ > $|\gamma_4^i|$ and $|\gamma_5^i| > |\gamma_6^i|$.

Aggregating as before over the sample of size *N*, the mean poll result is given, to a good approximation, by

$$\overline{Y}_t = \overline{y}_t + \overline{C}_t \tag{3.2}$$

where

$$\begin{aligned}\overline{C}_t = E(\gamma_1)D_t + E(\gamma_2)(1-D_t) + E(\gamma_3)D_t T_t + E(\gamma_4)(1-D_t)T_t \\ + E(\gamma_5)D_t T_t^2 + E(\gamma_6)(1-D_t)T_t^2\end{aligned} \tag{3.3}$$

where $E(\gamma_k)$ denotes the mean of the distribution of the γ_k^i coefficients in the voting population. Note, there is no need to require the α^i and the γ_k^i to be distributed independently in the population. It is of course possible that a mere shift of location is not sufficient to capture the 'in power' effect. The parameters β_j^i may have different distributions depending on whether the party is in power or in opposition, and since $\overline{y}_t$ is a weighted sum of different I(d) processes with weights $E(\beta_j)$, according to (2.6), its distribution would change accordingly. However, note that the value of d itself depends only on the distribution of α^i. There is no reason to suppose that this depends on whether the party in question is in power or in opposition.

We fit the election cycle to the series in Figs. 1(a) and 2(a) by estimating the regression defined by equations (3.2) and (3.3) by ordinary least squares, with $\log\{\overline{X}_t/(1-\overline{X}_t)\}$ replacing Y_t as response, and the component $\overline{y}_t$ treated as the residual. The parameter estimates are shown in Table 1, with reported standard errors in parentheses. Note that the predicted signs and relationships between the parameters hold in each particular. The predicted values and residuals from these regressions, our estimates of the series $\overline{C}_t$ and $\overline{y}_t$ respectively, are shown in Figures 1(b) and 2(b), and 1(c) and 2(c). Note how the periods spent in power by each party are clearly revealed in Figures 1(b) and 2(b), by the reversals of the cycle. The effect, as measured by the regression R^2, is more pronounced in the case of the Conservatives because of their longer tenure of office over the sample period.

	Conservative	Labour
$E(\gamma_1)$	-0.187 (0.0385)	-0.0251 (0.0577)
$E(\gamma_2)$	-0.532 (0.0476)	-0.376 (0.0468)
$E(\gamma_3)$	-0.0329 (0.0032)	-0.0366 (0.0059)
$E(\gamma_4)$	0.0284 (0.0048)	0.00250 (0.0039)
$E(\gamma_5)$	0.00055 (0.00006)	0.000632 (0.00012)
$E(\gamma_6)$	-0.0004 (0.00009)	-0.000038 (0.00007)
	R^2 =0.462	R^2 =0.0966

Table 1. Election cycle parameters

Least squares estimation of the election cycle is not a wholly satisfactory procedure, for two reasons. First, the variables $\bar{y}_t$, D_t and T_t are in principle jointly determined, in the sense that the switch points of D_t and T_t are potentially predictable from $\bar{y}_t$. This presents an estimation problem that would be very difficult to handle optimally. However, the number of switch points is small relative to the number of data points, and there are no grounds for fearing that the average correlation between $\bar{y}_t$ and the dummy explanatory variables is particularly large. Second, there is inefficiency due to the autocorrelation in $\bar{y}_t$ (see Beran, 1992, Section 2). While this problem could in principle be overcome by maximum likelihood estimation of the complete model, this refinement cannot be implemented with the existing software. However, the results reported below suggest that this would not make a large difference to the estimates of *d*.

4. Fractional Integration in the Gallup Data.

Our next step is to test both the raw series and the detrended series (in other words, the residual after removal of the election cycle) for long memory. We apply the modification due to Lo (1991) of the Hurst-Mandelbrot rescaled range (R/S) test (see Hurst (1951), Mandelbrot (1972), and Mandelbrot and Wallis (1968), among other references). Applied to a time series $X_1,...,X_T$, this is a test of the null hypothesis $d = 0$, based on the statistic

$$\frac{R}{S} = \frac{1}{\sqrt{T}\hat{\sigma}_n(q)}\left[\max_{1\le k\le T}\sum_{j=1}^{k}(X_j - \bar{X}_T) - \min_{1\le k\le T}\sum_{j=1}^{k}(X_j - \bar{X}_T)\right] \tag{4.1}$$

where $\bar{X}_T$ is the sample mean and

$$\hat{\sigma}_T^2(q) = \hat{\sigma}_X^2 + 2\textstyle\sum_{j=1}^{q}\omega_j(q)\hat{\gamma}_j\,,\ \ \omega_j(q) = 1 - j/(q+1) \tag{4.2}$$

where $\hat{\sigma}_X^2$ and $\hat{\gamma}_j$ are the usual estimates of the variance and *j*th autocovariance of the series, and *q* is an integer to be chosen. Equation (4.2) defines the estimator of the variance of the partial sum suggested by Newey and West (1987), which is consistent in the presence of weak dependence in the series, under the null hypothesis. When $d = 0$ the R/S statistic converges to a known limit distribution as *T* increases, but if $d > 0$ it diverges to ∞ and if $d < 0$ it converges in probability to 0.

We perform one-tail tests of two null hypotheses, $d = 0$ against the alternative $d > 0$, and $d = 1$ against the alternative of $d < 1$. In the second case, in effect a test of a unit root, this is done by differencing the data and using the lower tail of the null distribution. The values of the test statistics are shown in Table 2, for three different values of the lag truncation *q*. Given that *q* = 10 is almost certainly excessive according to the usual guidelines for the use of the Newey-West variance estimator (and also in view of our subsequent findings concerning weak dependence in these series) these results serve to confirm that we are dealing with true fractional processes. The asterisks denote rejection of the null at the 5% level, the critical values (from Lo 1991, Table II) being 0.861 for the lower 5% tail and 1.747 for the upper 5% tail.

Null hypothesis	$d = 0$			$d = 1$		
truncation lag	0	5	10	0	5	10
Conservative (raw data)	4.99*	2.19*	1.69	0.713*	0.980	1.057
Conservative (de-trended data)	3.77*	1.76*	1.40	0.452*	0.650*	0.772*
Labour (raw data)	5.26*	2.34*	1.81*	0.549*	0.792*	0.856*
Labour (de-trended data)	5.67*	2.55*	1.98*	0.587*	0.859*	0.966

Table 2. Tests for integer values of *d*.

Next, consider the estimation of *d*. To evaluate the robustness of the estimation to our assumptions, we consider a number of different cases. First, we estimate *d* both from the detrended series and the ‘raw’ series, to see how far our treatment of the election cycle affects the outcome. Second, we estimate it for the sub-period of continuous Conservative rule from June 1979 to May 1995 (192 observations) to check the adequacy of our treatment of the ‘in power’ effect by a shift dummy. Third, we do the estimation for quarterly data, taking every third observation. This last exercise is undertaken in the light of the result of Chambers (1994, Proposition 2), who shows that the value *d* that characterises a long-memory series is invariant to time aggregation. Of course, such an invariance property would not hold for a simple ARMA process.

We implemented two different estimation methods for *d*, a variant of the log-periodogram regression proposed by Geweke and Porter-Hudak (1983), and the maximum likelihood (ML) estimator implemented by Sowell (1992). Similar results were obtained in each case. We also performed Monte Carlo experiments with the log-periodogram regression, generating the data according to the Granger aggregation scheme, and these yielded the expected results. Here we report, in Table 3, only the ML estimates. (Details of the other analyses are available on request.) Note that the ML estimator is valid only for the stationary invertible case where $-0.5 < d < 0.5$, and the estimates in Table 3 are computed by adding 1 to the estimate of *d* obtained with the differenced data.

The relative robustness of these estimates to the removal of the election cycle is remarkable, noting how this feature dominates the Conservative data in particular. It is also interesting how closely the values of d from the Conservative and Labour series correspond. Note that $d \geq 1/2$ implies that the processes are nonstationary, although they are mean reverting. The ML estimator jointly estimates the parameters of the full ARFIMA(p,d,q) process, where p and q are given, and these have been chosen according to the Schwarz information criterion (SIC). Table 4 reports values of the SIC obtained for four estimated models. Interestingly enough, $p = 0$ and $q = 0$ are the best choices on this criterion, and the estimates in Table 3 correspond to these choices. On this basis, our series appear indistinguishable from ARFIMA(0,d,0) processes.

	Conservatives	Labour
Detrended data, Sept. 60-May 95	0.694 (0.040)	0.703 (0.037)
Raw Data, Sept. 60-May 95	0.779 (0.039)	0.726 (0.037)
Detrended data, June 79-May 95	0.765 (0.063)	0.717 (0.055)
Quarterly data 60Q3-95Q2	0.601 (0.066)	0.693 (0.069)

Table 3. ML Estimates of d (standard errors in parentheses).

Model		Conservatives		Labour	
p	q	Sept. 60-May 95	June 79-May 95	Sept. 60-May 95	June 79-May 95
0	0	–595	–265	–523	–218
1	0	–590	–261	–522	–216
0	1	–589	–261	–520	–212
1	1	–584	–256	–517	–215

Table 4. SIC for ML estimation of ARFIMA(p,d,q) in the detrended data.

The evidence for pure fractional processes revealed by Table 4 is intriguing, although as noted above, we cannot deduce anything directly from it about the characteristics of the W_{jt} processes. The filter that the aggregation applies to these processes in our model, according to (2.6), depends upon both parameters of the Beta distribution, as indicated by (2.8). The value of v determines the remote behaviour of the moving average coefficients, but the low-order coefficients also depend on the value of u. If we knew that the W_{jt} were white noise processes, we could deduce an approximate value for u in (2.8), once v has been determined from d, by matching up the coefficients in (2.8) and (2.10). Alternatively, knowledge of u would let us deduce whether or not the W_{jt} were white noise processes.

Lacking either piece of information, we have an identification problem. However, if we provisionally accept the white-noise postulate, we can tentatively use our pure fractional noise finding for the poll series to deduce a plausible value for u. Consider the fact (see e.g. Granger 1980, p. 229) that in the autoregressive representation of the pure fractional noise,

$$(1-L)^d x_t = u_t \, , \tag{4.3}$$

the lag polynomial $a(L) = (1-L)^d$ has coefficients

$$a_k = \frac{\Gamma(k-d)}{\Gamma(-d)\Gamma(k+1)}, \; k = 0,1,2,... \tag{4.4}$$

Assuming that the W_{jt} processes are white noise, we can therefore reconstruct the intertemporal structure of the time series process

$$u_t = (1-L)^d \, \bar{y}_t = \left(\sum_{j=0}^{\infty} a_k L^k \right) \left(\sum_{j=0}^{\infty} \alpha_k L^k \right) \bar{\varepsilon}_t \tag{4.5}$$

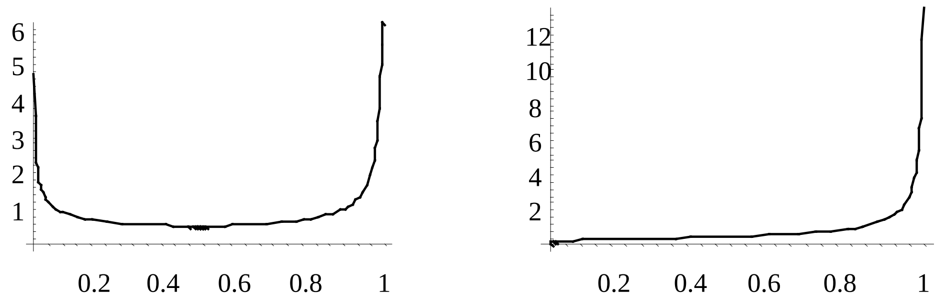
where the α_k are the coefficients in (2.8), and

$$\bar{\varepsilon}_t \simeq \sum_{j=1}^{J} E(\beta_j) W_{jt} \, . \tag{4.6}$$

Table 5 lists, for the case v = 0.3 and three different cases of u, the first 10 coefficients (that is to say, the coefficients of L^j for j = 0,...,9) of the product of polynomials appearing in the right-hand member of (4.5). Figure 4 shows the Beta(u,v) densities for two of these cases. The third (intermediate) case appears as Figure 3 above. Thus, with v set so that $1 - v$ lies near the estimated values of d, most of the ‘plausible’ cases of u, those which give neither too much nor too little weight to the values of α close to 0, arguably lie between the two cases in Figure 4. Assuming the sequence of moving average coefficients is monotone beyond some point, and making reasonable continuity assumptions, we take Table 5 to provide evidence that all these cases give rise to aggregate processes which resemble pure fractional noise fairly closely. In other words, if our model is correct, the finding of ARFIMA(0,d,0) processes is a predictable and robust implication of white noise news processes.

lag	u = 0.6	u = 0.45	u = 0.3
0	1.	1.	1.
1	0.0694	0.00852	-0.091
2	0.0231	-0.00097	-0.031
3	0.0116	-0.0014	-0.016
4	0.0070	-0.0012	-0.010
5	0.0048	-0.00096	-0.0071
6	0.0026	-0.00078	-0.0052
7	0.0021	-0.00064	-0.0040
8	0.0017	-0.00053	-0.0032
9	0.0014	-0.00045	-0.0026

Table 5. Coefficients of $(1-L)^d \alpha(L)$



(a) u = 0.3, v = 0.3 (b) u = 0.6, v = 0.3

Fig. 4. Beta(u,v) densities

5. Analysis of the Fractionally Differenced Series

The series shown in Figures 1(d) and 2(d) are the results of applying the fractional-difference filter $(1-L)^d$ to the detrended series in 1(c) and 2(c), where d is the ML estimate of the differencing parameter from the first row of Table 3. These series have a correlation coefficient of –0.46, showing that the fortunes of the parties are linked closely although not rigidly. The Liberal/Liberal Democratic party and the nationalist parties act as a buffer between the main contenders, and are often perceived to absorb shifts in support, potentially from either side although more often at the expense of Labour.

To analyse these series further, our first step is to to examine the evidence that they are not merely stationary but white noise, using conventional time series methods. Table 6 gives the relevant statistics, in addition to the Jarque-Bera (1980) tests for normality, which is strongly rejected in both cases. All of these statistics are asymptotically chi-squared with degrees of freedom indicated under the null hypothesis, and asterisks denote rejection at the nominal 5% level. The Box-Pierce Q statistics were actually calculated for all orders of lag up to $n/3$, but none of those not quoted here were significant at the 5% level. The evidence for time-dependence is clearly weak, particularly in the Conservative series, which shows no linear correlation at all, and just a possibility of autoregressive conditional heteroscedasticity (or ARCH, see Engle 1982). To the extent that the defining property of 'white noise' is the absence of autocorrelation, the absence of ARCH effects indicates a stronger restriction on dependence. Tests for ARCH are known to have power against general forms of nonlinear time dependence (see for example Brock, Hsieh and LeBaron 1991, page 60). The Labour series yields evidence of autocorrelation at second order, but none at first order, nor at any higher order. This pattern is difficult to account for, and is conceivably just an outlier in the null distribution of the statistic. Some mild ARCH effects are also found here, after removing the autocorrelation in mean by a second order autoregression. On the whole, these findings lend support to the explanation put forward at the end of Section 4.

Test	d.f.	Conservative	Labour
Jarque-Bera	2	24.58*	30.79*
Box-Pierce Q	1	0.131	1.51
	2	0.306	8.32*
	12	17.45	19.64
ARCH	1	4.04*	4.49*
	2	5.23	7.76*
	12	9.11	20.12*

Table 6. Statistical properties of the fractionally differenced series

Our second analysis, which from the perspective of a theoretical explanation of voting behaviour is the most critical one, seeks correlations between the fractionally differenced series and economic indicators. An important implication of the finding of nonstationarity of the opinion series is that it calls the results of earlier regression studies into question. It has been shown (see Tsay and Chung 1995) that fractional processes can share with I(1) processes a liability to the well-known 'spurious regressions' phenomenon first demonstrated by Granger and Newbold (1974). To construct a test for economic influences that avoids such pitfalls, the natural approach is to model the news process driving support, rather than support itself. In other words, we should attempt to estimate equation (4.6), by regressing the fractional differences u_t, which we assume approximate the $\bar{\varepsilon}_t$ according to (4.5), on to as many candidate W_{jt} processes as we are able to observe. The fact that our estimates of the difference processes are found to be close to serially independent without further filtering is strongly suggestive of the view that these processes represent 'news', and so $\bar{\varepsilon}_t$ should be explained by surprises in economic indicators.

To this end, we use monthly survey data on the anticipations of brokers and other financial market professionals of growth in the Retail Prices Index, average earnings, and industrial production. Our measures of surprises are the differences between the actual out-turns and the sample median predictions of these growth rates. In addition to these variables, we used the actual monthly change in the mortgage rate (the average of the Abbey National and the Halifax building societies' published rates). Regressing the fractionally-differenced series on to these four variables lagged one period (to allow for synchronicity of the survey dates), for the period December 1982-June 1994, yielded the results shown in Table 7.

	Conservative		Labour	
	Coeff.	*t* ratio	Coeff.	*t* ratio
Intercept	-0.0173	-1.643	0.0115	0.992
RPI surprise	0.0844	2.0092	-0.000065	-0.0014
Ave. earnings surprise	0.00872	0.213	-0.0338	-0.748
Growth surprise	0.00927	1.030	-0.0177	-1.782
Rise in mortgage rate	-0.0220	-1.024	-0.0121	-0.511
	R^2 = 0.0437		R^2 = 0.0301	

Table 7. Regression analysis of the fractionally differenced series

Consider these regression coefficients as estimates of certain of the parameters $E(\beta_j)$. Since the specification is incomplete we cannot rule out the possibility of bias, but the sample period is one of continuous Conservative rule, so it is reasonable to assume parameter stability. The effects are mostly weak, although most of the signs accord with our expectations. However, note that these coefficients will be large only when the news is important *and* there is a consensus among voters about its meaning.

It is possible to have $E(\beta_j) \approx 0$ even when the population includes large absolute values of the β_j, if these differ in sign. The most interesting discovery is that inflation surprises have a significant positive effect on Conservative support. The fact that voters like inflation should not surprise us very much, however, given that inflation erodes debt, and can have distributional effects which are advantageous to families.

Finally, in Table 8, we compare the outlying fractional differences, corresponding to months of major swings in support, with the diary of national and international political events. The outliers in question are labelled in Figures 1(d) and 2(d), and the final column of the table shows the size of the difference expressed in standard deviations. There are however a number of outliers in the series that cannot be easily accounted for in this way. For example, the single striking Times headline in the month of one large Conservative dip (August 1985) turns out to be "Poll Shock for Thatcher"! This also accords with our view that what drives opinion is the aggregate of a large number of news items whose significance is different for different voters, and hence have a small weight in the aggregate. The larger shifts in opinion take place, according to this view, when events occur whose significance voters happen to agree on, although they need not be in themselves of particular weight.

Key	Date	Event	SDs
Conservatives (see Fig. 1d)			
(1)	February 1975	Thatcher replaces Heath	+3.6
(2)	May 1982	Falklands war	+3.8
(3)	October 1984	Brighton bomb	+2.8
(4)	March 1990	Introduction of the Poll Tax	–2.2
(5)	December 1990	Major replaces Thatcher	+2.5
Labour (see Fig. 2d)			
(6)	February 1979	‘Winter of Discontent’	–3.4
(7)	February 1981	SDP launched	–3.5
(8)	October 1981	Benn nearly wins deputy leadership	–3.3
(9)	October 1983	Kinnock replaces Foot	+3.6

Table 8. Dates of some major opinion shifts

6. Conclusion

This paper presents a model of the evolution of political support that is able to account for the striking observation that the Gallup poll series for Conservative and Labour support in the UK can be modelled as pure fractional noise processes. (Since the logistic transformation is approximately linear in the range of the data points, this generalisation is valid in practice for both the original and the transformed series.) The conclusions of our study are as follows. First, voters react to news of economic and political events, but not always in a manner that is predictable in the aggregate. Second, the average voter forgets eventually, but not rapidly. Voters have a long memory of events. Party loyalties are persistent although, on average, not so persistent as to rule out changes of allegiance in the long run. One of the particularly interesting findings is the stability in the value of d (and hence of v) both across the parties and across sub-periods of the data. This provokes the speculation, which we hope to test in future research with different data sets, that the distribution of the attribute of political commitment across the voting population might be a relatively stable sociometric constant, and does not depend upon variable factors such as left-right orientation, class structure, trade union affiliation, and the like.

The current (1992-96) period in British politics is of particular interest, in that the swing against the Conservatives during the Major administration is of a magnitude unprecedented in the history of political polling. Predicting the period therefore presents a comparatively severe test of our model. We first undertake an *ex post* forecasting exercise by predicting, from the standpoint of May 1995 when our study began, the seventeen extra observations that have since become available to us.

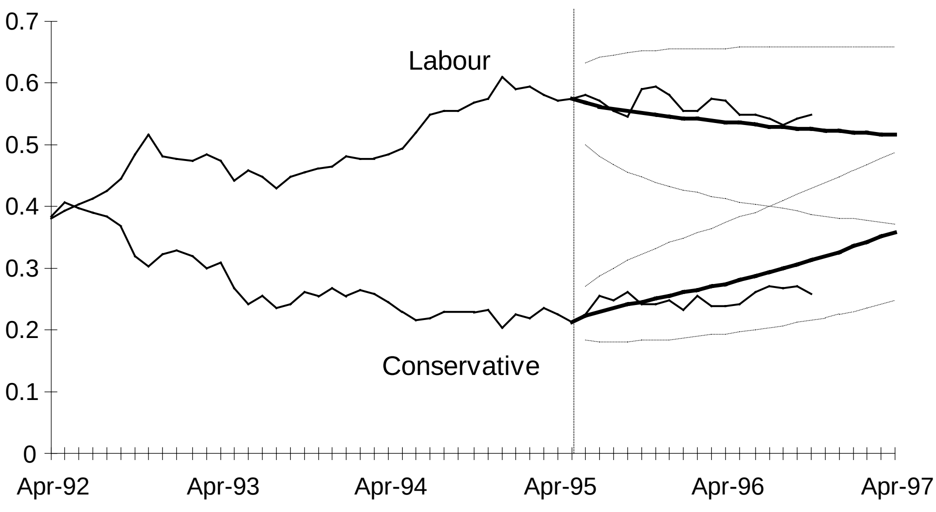


Fig. 5 Multi-step forecasts from May 1995, with out-turns to October 1996

Fig. 5 shows the actually recorded percentage support for Conservative and Labour for the period, with the vertical broken line marking the limit of the data used to fit the model. The multi-step model forecasts for the period June 1995 to April 1997 (the limit of the government's period of tenure) are shown as heavy lines, with 2-standard error bands shown as dotted lines. The chart shows that our model continues to track the actual trends in support fairly well, even 17 months ahead. The mean forecast errors are −1.6% for the Conservatives, and +1.8% for Labour, with root mean squared errors of 2.7% and 2.4% respectively.

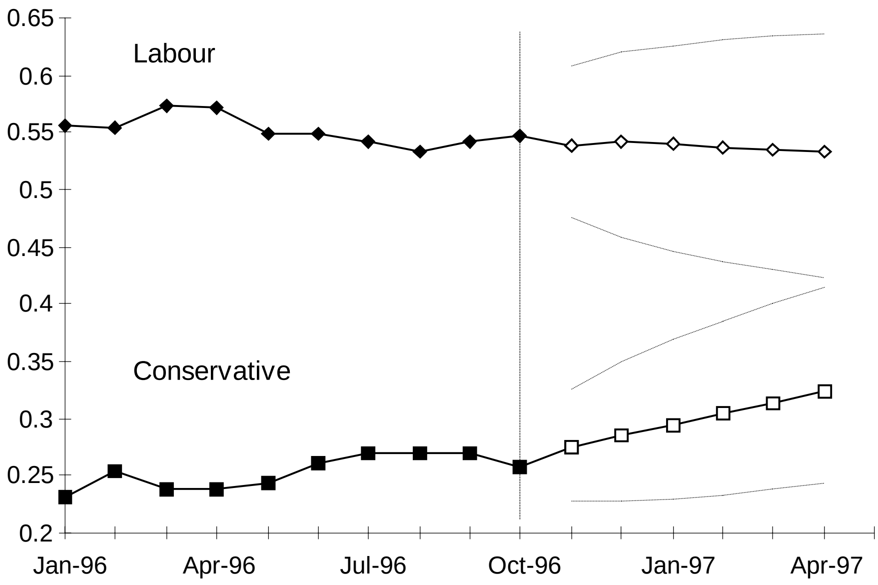


Figure 6: Multi-step forecasts, November 1996 to April 1997

It is of obvious interest to use our model to project the outcome of the 1997 election, which lies up to six months in the future at the time of writing. Fig. 6 shows, similarly to Fig. 5, the *ex ante* forecasts for November 1996 to April 1997 using the additional observations up to October 1996, although the same parameter estimates as before. The percentage support forecast in election month is 32.3% for the Conservatives with an approximate standard error of 4.2%, and 53.3% for Labour with an approximate standard error of 5.3%. We note that there was a shift in Conservative support from 31.5% to 45% over the two months April-June 1982, the period of the Falklands War. It would evidently require a shock of similar magnitude to shift the parties back to their 1992 positions in time for the election. We do not overlook the fact that this is a forecast of poll results, not of the election result itself, and the chastening experience of the polling organisations at the 1992 election (see Smith 1996 for an analysis) reminds us that these things can be different. We have not addressed the problem of poll bias in this paper, and it remains unclear whether or not this will prove a transient phenomenon. Subject to this caveat, the evidence appears strong enough for us to forecast a Labour victory with some confidence.

Acknowledgements

We thank a referee and the editor for helpful drafting suggestions, and additional references. We are also grateful to Emily-Jane Raeburn and Jasbir Sandhu for assistance with the computing, to the Gallup Organisation for supplying the opinion poll data, to Money Market Services for the expectations of economic indicators, and to Fallaw Sowell for providing his ARFIMA estimation package. The research was partly supported by a grant from Inquire.

The data used in this study are available from the authors on request.

References

Beran, Jan (1992) 'Statistical methods for data with long-range dependence', *Statistical Science* 7 (4) 404-427

Byers, J.D.(1991) 'Is political popularity a random walk?', *Applied Economics* 23, 967-974.

Brock, William A., David A. Hsieh and Blake LeBaron (1991) *Nonlinear dynamics, chaos and instability: statistical theory and economic evidence*, Cambridge, Mass: MIT Press

Chambers, M. (1994) 'Long memory and aggregation in macroeconomic time series' University of Essex, Department of Economics Working Paper No. 437.

Chrystal, K.A. and D. Peel (1987) 'What can economics learn from political science and vice versa?' *American Economic Review* (Papers and Proceedings) 76, 2, 62-65

Davidson, J. (1994) *Stochastic Limit Theory*, Oxford University Press

Diebold, F. X. and G. D. Rudebusch (1989) 'Long memory and persistence in aggregate output' *Journal of Monetary Economics* 24, 189-209

Engle, R. F. (1982) 'Autoregressive conditional heteroscedasticity, with estimates of the variance of United Kingdom inflation' *Econometrica* 50, 987-1007

Frey, B. S. and F. Schneider (1978) 'A politico-economic model of the United Kingdom', *Economic Journal* 88, 243-253

Gärtner, M. (1994) 'Democracy, elections and macroeconomic policy: two decades of progress', *European Journal of Political Economy* 10, 85-109

Geweke, J. and S. Porter-Hudak (1983) 'The estimation and application of long memory time series models', *Journal of Time Series Analysis* 4, 221-238

Goodhart, C. A. E. and R. J. Bhansali (1970), ‘Political economy’, *Political Studies* 18, 265-279

Granger, C. W. J. (1980) ‘Long memory relationships and the aggregation of dynamic models’, *Journal of Econometrics* 14, 227-238

Granger, C. W. J. and R. Joyeux (1980) ‘An introduction to long-memory time series models and fractional differencing’, *Journal of Time Series Analysis* 1,1, 15-29

Granger, C. W. J. and P. Newbold (1974) ‘Spurious regressions in econometrics’, *Journal of Econometrics* 2, 111-120

Hall (1978) ‘Stochastic implications of the life-cycle-permanent income hypothesis: theory and evidence’, *Journal of Political Economy* 86, 971-987

Haslett, John and Adrian E. Raftery (1989) ‘Space-time modelling with long-memory dependence: assessing Ireland’s wind power resource’ *Applied Statistics* 38 (1) 1-50

Holden, K. and D. Peel (1985) ‘An alternative approach to explaining political popularity’, *Electoral Studies* 4, 231-239

Hosking, J. R. M. (1981) ‘Fractional differencing’, *Biometrika* 68, 165-176.

Hurst, H. (1951) ‘Long term storage capacity of reservoirs’, *Transactions of the American Society of Civil Engineers* 116, 770-799

Jarque, C. M. and A. K. Bera (1980) ‘Efficient tests for normality, homoscedasticity and serial independence of regression residuals’, *Economics Letters* 6, 255-259

Lo, Andrew W. (1991) ‘Long-term memory in stock market prices’, *Econometrica* 59, 5, 1279-1313

Mandelbrot, B. (1972) ‘Statistical methodology for non-periodic cycles: from the covariance to R/S analysis’, *Annals of Economic and Social Measurement* 1, 259-290

Mandelbrot, B. and J. Wallis (1968) ‘Noah, Joseph and Operational Hydrology’ *Water Resources Research* 4, 909-918

Miller, W. L. and M. Mackie (1973) ‘The electoral cycle and the asymmetry of government and opposition popularity: an alternative model of the relationship between economic conditions and political popularity’, *Political Studies* 21, 265-279

Minford, A. P. L. and D. A. Peel (1982), ‘The political theory of the business cycle’, *European Economic Review* 17, 253-270

Newey, W. and K. West (1987) ‘A simple positive definite heteroscedasticity and autocorrelation consistent covariance matrix’, *Econometrica* 55, 703-705

Pissarides, C. (1980) ‘British government popularity and economic performance’, *Economic Journal* 90, 569-81

Rogoff, K. and A. Sibert (1989) ‘Equilibrium political business cycles’, *Review of Economic Studies* 55, 1-16

Schneider, F. and B. S. Frey (1988) ‘Politico-economic models of macroeconomic policy: a review of the empirical evidence’ in T. D. Willett (ed.) *Political Business*

Cycles: The Political Economy of Money, Inflation and Unemployment (Duke University Press, Durham N.C.) 239-275

Scott, A. J., T. M. F. Smith and R. G. Jones (1977) 'The application of time series methods to the analysis of repeated surveys', *International Statistical Review* 45, 13-28

Smith, T. M. F. (1996) 'Public opinion polls: the UK general election 1992', *Journal of the Royal Statistical Society A*, 159 (3), 535-545

Sowell, F. (1992) 'Maximum likelihood estimation of stationary univariate fractionally integrated time series models', *Journal of Econometrics* 53, 165-188

Tsay, W-J. and C-F. Chung (1995) 'The spurious regression of fractionally integrated processes', Econometrics and Economic Theory Paper No. 9503, Michigan State University (September).