

JOURNAL OF APPLIED ECONOMETRICS

J. Appl. Econ. **18**: 137–155 (2003)

Published online 14 January 2003 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jae.700

LONG-RANGE DEPENDENCE IN SPANISH POLITICAL OPINION POLL SERIES

JUAN J. DOLADO,^a JESÚS GONZALO^{b*} AND LAURA MAYORAL^c

^a *Dept of Economics, Universidad Carlos III de Madrid, Spain*

^b *Dept of Statistics and Econometrics, Universidad Carlos III de Madrid, Spain*

^c *Dept of Economics and Business, Universidad Pompeu Fabra, Spain*

SUMMARY

This paper investigates the time series properties of partisanship for five political parties in Spain. It is found that pure fractional processes with a degree of integration, d , between 0.6 and 0.8 fit the time-series behaviour of aggregate opinion polls for mainstream parties quite well, whereas values of d in the range of 0.3 to 0.6 are obtained for opinion polls related to smaller regional parties. Those results are in agreement with theories of political allegiance based on aggregation of heterogeneous voters with different degrees of commitment and pragmatism. Further, those models are found to be useful in forecasting the results of the last general elections in Spain. As a further contribution, new econometric techniques for estimation and testing of ARFIMA model are used to provide the previous evidence. Copyright © 2003 John Wiley & Sons, Ltd.

1. INTRODUCTION

Over the last few years there has been a growing interest among political scientists in applying time series techniques to analyse the statistical properties of aggregate political popularity data in various formats such as approval levels and partisanship measures. For example, Box-Steffensmeir and Smith (1996), Byers *et al.* (1997), Eisinga *et al.* (1999) and various articles included in a recent special issue of *Electoral Studies* (2000) report evidence for the United States, the United Kingdom and several other OECD countries, indicating that the time series of poll ratings in those countries are well modelled by fractionally-integrated processes which present high persistence but that eventually revert to their mean. Those findings differ from the statistical implications of previous theories of political support which, upon the standard assumption that voters form rationally their expectations about future events, argue that the effect on voting intentions of news about the economy or the political environment is permanent. Statistically, this implies the presence of a unit root in the process governing the time series of political opinion data; cf. Holden and Peel (1985) and Chrystal and Peel (1987). By contrast, the above evidence on long-memory behaviour supports theories of political allegiance based on the aggregation of heterogeneous voters with different attributes of commitment and pragmatism which help to provide explicit microfoundations of aggregate-level measures of partisanship as a function of individual-level opinions and attitudes (see Achen, 1975).

* Correspondence to: Jesús Gonzalo, Department of Statistics and Econometrics, Universidad Carlos III de Madrid, 126 28903 Getafe, Madrid, Spain. E-mail: jgonzalo@est-econ.uc3m.es
Contract/grant sponsor: DGESIC; Contract/grant number: SEC01-0890.

Within the extensive literature on the dynamics of aggregate partisanship, two papers have been especially instrumental for our own research. First, Byers, Davidson and Peel (1997) (BDP henceforth) present a model to characterize the time-series behaviour of aggregate opinion polls for the Conservative and Labour parties in the UK. Their model links partisanship at the individual and aggregate level by classifying voters into two broad categories: 'committed' voters (either insensitive to news or sluggish to change allegiance in the face of news) and 'floating' voters (sensitive to news and performance of the different political parties). Assuming a certain type of distribution for the attributes pertaining to both types of voters, they make use of a statistical result by Granger (1980) on the aggregation of heterogenous stationary first-order autoregressive (AR(1)) processes to achieve long-memory and mean reversion in the time-series behaviour of partisanship at the aggregate level, reconciling in this way the above-mentioned empirical findings with the theory. Next, there is the work by Box-Steffensmeier and Smith (1996) (henceforth BSS) who use a slightly different model for individual partisanship due to Franklin and Jackson (1983) (FJ henceforth), in combination with the previous aggregation result, to obtain further evidence in favour of long memory in the time-series behaviour of aggregate partisanship for Republicans and Democrats in the USA.

In this paper, we use a slightly modified version of the intersection of the two previous approaches to examine the extent to which the earlier results apply to Spanish poll data. The use of this data set is particularly interesting since it includes information on the voting intentions of both mainstream parties, like in the previous studies, and smaller regional parties whose political support may behave according to a different pattern. To deal with this diversity of political parties, our approach starts by allowing for the existence of two types of voter: those who hardly react to 'news' ('militants') and those who react more often ('non-militants'). It is within the latter group that we follow BDP in distinguishing between 'committed' and 'floating' voters, according to the degree of sluggishness in their voting intentions. The distinction between 'militants' and 'non-militants' is an important one since, as will be seen below, the small regional parties in Spain have a voting intention which is fairly constant over time. In such a case, the time-series analysis of deviations of the partisanship series around their sample mean could give the wrong impression that a majority of voters for these parties belong to the 'floating' class, since persistence is low. However, in our view, a more appropriate interpretation would be that a large fraction of their electorate belongs to the 'militant' class, namely, that their voting intentions are governed by a non-zero-mean process with a small variance relative to the size of the unconditional mean.

Besides studying partisanship for small regional parties, the analysis of political support for Spanish mainstream parties is of some interest since Spain only regained democracy in 1977 after forty years of dictatorship in which political parties were illegal. In this sense, it should be remarked that the length of the available political support series is shorter than in other countries with a much longer and mature democratic tradition. Despite this shortcoming, however, it seems interesting to study whether the statistical properties of the opinion polls in Spain conform to the patterns found for other countries.

From the econometric viewpoint, a further contribution of our work is the application of newer and seemingly more robust techniques in finite samples to the estimation and testing of the degree of fractional differencing, d , in autoregressive fractionally integrated moving average (ARFIMA) models. As regards estimation, we use a new minimum-distance estimator of d proposed by Mayoral (2000) which, besides being computationally more tractable than most of the available methods in the literature, can be used to implement a very simple test of the null hypothesis that d is equal to unity versus the alternative that d is less than unity that, according to the previous

discussion, are the two hypotheses of interest in the theory of political popularity. This test, which shares the spirit of the well-known Dickey–Fuller test for unit roots, has been recently developed by Dolado, Gonzalo and Mayoral (2002) (DGM henceforth) in a companion paper. It turns out to be easily implementable in the time domain and has both better size and power properties in finite samples than many of the tests which are typically used in applied work.

The rest of the paper is structured as follows. Section 2 summarizes the microfoundations of the model of voting intentions proposed by BDP, using the model for partisanship proposed by FJ. In this way we emphasize the key assumptions which give rise to an ARFIMA process as an appropriate model governing the time-series behaviour of aggregate poll series. Section 3 explains both the estimation and testing approaches used in this paper. In Section 4, we report the results concerning the voting intentions of Spanish voters about the five main political parties which have remained active during the period 1978–99, out of which three are mainstream parties and two are regional parties. We discuss and interpret our key empirical findings in the light of the taxonomy of voters described above and compare our results to those found for other countries with a longer democratic tradition. Some forecasting results, relevant for the last general elections held in March 2000, which were carried out before it took place, are also reported. Finally, Section 5 concludes.

2. A PARTISANSHIP-POLITICAL POPULARITY MODEL

Until recently, the conventional approach to modelling the time series properties of aggregate political popularity was based on the following considerations (see e.g. Chrystal and Peel, 1987). Let \bar{X}_t be the average propensity to support a given party which is assumed to depend linearly on the expected present value of the benefits (V_t) that the supporters would receive in case the party wins the elections, i.e. $\bar{X}_t = \varphi V_t + e_t$, where e_t is an i.i.d. error term. If expectations are formed rationally, then it is well known (as in the case of the Permanent-Income Hypothesis of consumption) that the change in V_t , ΔV_t , is an innovation, ξ_t , namely, that the best linear predictor of V_t is V_{t-1} . Thus, $\Delta \bar{X}_t = \varphi \xi_t + \Delta e_t$, implying that \bar{X}_t is an $I(1)$ process with a $MA(1)$ error term. The idea behind both the BDP and BSS approaches is to relax the $I(1)$ characterization of \bar{X}_t to the more general class of fractionally-integrated, $FI(d)$, processes, where d is fractional number rather than an integer one. They do so by assuming that there is a continuum of voting intentions that are captured by AR(1) processes with different degrees of persistence. Despite the existence of a continuum of voting intentions, it will be useful in the following discussion to focus on those which have large, yet stationary, autoregressive roots and on those which have low roots. The former pertain to ‘committed’ voters whilst the latter belong to the ‘floating’ type. Then, application of an aggregation argument due to Granger for heterogenous AR(1) processes leads to long memory in the aggregate poll series.

In our opinion, however, it is preferable to depart slightly from the previous analysis by defining an initial taxonomy of voters in terms of two broader categories, labelled as ‘militants’ and ‘non-militants’, where the latter group includes the above-mentioned types of ‘committed’ and ‘floating’ voters. Hence, overall, there will be three groups of voters. ‘Militants’ are those voters with a strong allegiance to a particular party such that their expected support for that party, conditional on past actions, is basically time invariant. By contrast, ‘non-militant’ voters do not have such a strong identification with a specific political party. Therefore, they tend to change their expected support over time in reaction to ‘news’ affecting the potential performance of the party, albeit in different ways. ‘Committed’ voters weigh the historical context almost equally to current performance and

therefore possess a long memory of the performance of all parties. ‘Floating’ voters, in turn, are more pragmatic in outlook and discount the past fairly quickly. Moreover, their future voting behaviour is typically more difficult to forecast than that of ‘committed voters’. Thus, the degree of persistence on voting intentions at the aggregate level will hinge upon the distribution of these types of voters in the total voting population.

To motivate the previous taxonomy of voters, it is convenient to assume, as is customary in the political science literature (see FJ), that the individual partisanship (or ‘party identification’), namely the strength of support for a given party, follows a dynamic model which is captured by an AR(1) process, possibly with a drift. The different types of voters could be characterized in terms of the sizes of the AR parameter and the drift. Accordingly, a political party with a large share of ‘militants’ will be one in which most individual series present a large drift relative to the dispersion of the series and small persistence in the AR(1) process. In turn, a party with a large share of ‘committed’ voters implies that the individual series will have both high values of the AR root and a small drift. Finally, a party with a large proportion of ‘floating’ voters implies both a low value of the AR root and a small drift in the individual series.

More specifically, following BSS, we define an unobservable latent variable Y_t^i as the i th individual’s partisanship at time t . That variable can be thought of as being the sum of two components, i.e. $Y_t^i = C_t + y_t^i$. On the one hand, C_t can be interpreted as a common time-varying component which is deemed to capture the effect of some aggregate events like unemployment or inflation. Further, it can capture the effects of the ‘election cycle’ since, in some countries (particularly those with a long democratic history like the UK or the USA), there is evidence that shifts in popularity tend to depend on whether a party is in power or in opposition, and also on the timing of the poll, namely, whether it is close or far away from an election. On the other hand, the unobserved latent variable y_t^i can be interpreted as the stochastic individual component of partisanship. According to FJ, it can be defined as a continuous unidimensional variable representable in the real number line, with the zero point representing indifference between supporting a given party, (say) A , and the complement of A , (say) state B (i.e. voting for another party or abstention). Increasing positive values of y_t^i represent monotonically increasing partisanship for A , whereas increasingly negative values indicate greater preferences for B , i.e. greater dislike for A . Following the discussion above, it is assumed that the stochastic component of party identification at time t , y_t^i , depends on last period’s partisanship, y_{t-1}^i , some individual fixed influence (c^i) and a random disturbance (ε_t^i), according to the stationary AR(1) specification

$$y_t^i = c^i + \alpha^i y_{t-1}^i + \varepsilon_t^i, \quad 0 \leq \alpha^i < 1 \quad (1)$$

where c^i and α^i are assumed to be independent. A high value of c^i together with both a low value of α^i and low variance of ε_t^i would be interpreted as being a ‘militant’, in the sense that, apart from some minor random shocks, partisanship is basically a constant. When such is not the case, i.e. for ‘non-militants’, a value of α^i smaller but close to 1 is attached to a ‘committed’ voter whereas a value close to 0 corresponds to a ‘floating’ voter. Thus, for the former type of voter, partisanship is basically predetermined by their past preferences whilst, for the latter, it changes more freely in the face of news. In sum, in the absence of large persistence, the relative size of c^i with respect to the variance of ε_t^i measures the degree of ‘militancy’ whereas, under persistence, the size of α^i captures the degree of ‘commitment’ so that a high value (low) is a sign of ‘committed’ (‘floating’) voting behaviour. At this stage, it should be once more remarked that although our previous discussion is framed in terms of three types of voters for heuristic reasons,

the statistical model underlying the properties of the aggregation of individual partisanship relies upon the existence of a continuum of types which are then aggregated according with some mixing distribution for the α^i 's.

Following BDP and BSS, in order to characterize the time-series behaviour of the aggregate polls, the next step is to apply Granger (1980)'s result for the cross-section aggregation of heterogenous AR(1) processes, i.e. $y_t = \sum_1^N y_t^i$. For that, it is assumed that the distribution of attributes of commitment and pragmatism in the voting population, represented by the α^i 's in (1), is continuous on the support $[0,1)$ in such a way that part of the probability mass is located near 1, but excluding it. Following Granger (1980), a class of distributions which accomplishes the required characteristics is the modified *Beta*(p, q) distribution

$$dF(\alpha) = 2 \frac{1}{B(p, q)} \alpha^{2(p-1)} (1 - \alpha^2)^{q-1} d\alpha \quad (2)$$

with $0 < q < 2$, obtained from the standard *Beta* distribution $dF(x)$ with the change of variable $x = \alpha^2$ and where $B(p, q)$ denotes the Beta function. Notice that the shape of the distribution is symmetric if $p = q$, convex if $p, q < 1$ and concave if $p, q > 1$. For alternative values of p and q inside the unit circle, values of α close to 1 will yield a sizeable probabilistic mass near the point 1 itself. It should be noted that the choice of mixing distribution is not restricted to the *Beta* density function, as Lippi and Zaffaroni (2000) have recently proved. According to these authors, the long-memory property of the aggregation of heterogenous stationary AR(1) processes still holds if the mixing distribution belongs to the family $B(\alpha, \theta) \sim C_\theta (1 - \alpha)^{b(\theta)}$, where ' \sim ' stands for asymptotic equivalence when α tends to 1^- and $b(\theta)$ is a function from $\Theta \subseteq \mathfrak{R}^s, s \geq 1$ to $(-1, \infty)$. A particular case of this general family of mixing distributions is the *Beta* distribution. In addition, these authors prove that the aggregate process remains fractionally integrated when the individual series follow stationary ARMA processes instead of pure AR processes.

Since the power spectrum of y_t^i is:

$$f_i(w) = \frac{1}{2\pi} \frac{\sigma_\varepsilon^2}{|1 - \alpha^i z|^2}, \quad \text{with } z = e^{-iw} \quad (3)$$

where $i = \sqrt{-1}$, then, assuming independent components and that the ε^i 's are also independent of the α^i 's, we can get the following approximation for the power spectrum of the sum of the y_t^i 's, i.e. $y_t = \sum_1^N y_t^i$, denoted as $f(w)$:

$$f(w) \simeq \frac{N}{2\pi} E \left(\text{var}(\varepsilon_t^i) \int \frac{1}{|1 - \alpha z|^2} dF(\alpha) \right) \quad (4)$$

Hence, noting that

$$\frac{1}{|1 - \alpha z|^2} = \frac{1}{(1 - \alpha)^2} \left[\frac{1 + \alpha z}{1 - \alpha z} + \frac{1 + \alpha \bar{z}}{1 - \alpha \bar{z}} \right] \quad (5)$$

where \bar{z} is the complex conjugate of z and using $(1 + \alpha z)/(1 - \alpha z) = 1 + 2 \sum_1^\infty (\alpha z)^i$, it follows from (4) and (5) that the coefficient of z^k in $f(w)$ is:

$$(2/B(p, q)) \int_0^1 \alpha^{2p+k-1} (1 - \alpha^2)^{q-2} d\alpha \quad (6)$$

Next, from the standard definition of the power spectrum, it follows that (6) is the k th autocovariance of y_t , namely $E(y_t, y_{t-k})$. Then it is easy to check that:

$$2\alpha^{2p+k-1}(1-\alpha^2)^{q-2} = \alpha^{2(p+k/2-1)}(1-\alpha^2)^{(q-1)-1}$$

which yields

$$E(y_t, y_{t-k}) = \frac{B(p+k/2, q-1)}{B(p, q)} = \frac{\Gamma(q-1)\Gamma(p+k/2)}{B(p, q)\Gamma(p+k/2+q-1)} \quad (7)$$

since $B(p, q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$ where $\Gamma(\cdot)$ is the gamma function. Using Sterling's theorem, namely $\Gamma(j+a)/\Gamma(j+b)$ is of order j^{a-b} , it follows that, for k sufficiently large,

$$E(y_t, y_{t-k}) = O(k^{1-q}) \quad (8)$$

It is well known (see Granger and Joyeux, 1980) that if x_t is a fractionally integrated process of order d , $FI(d)$, then,

$$E(x_t x_{t-k}) = O(k^{2d-1}) \quad (9)$$

Therefore, comparison of (8) and (9) yields that $y_t \sim FI(1-q/2)$. This implies that y_t will be a non-stationary, albeit mean-reverting, process if $0 < q \leq 1$ (i.e. $1 > d \geq 1/2$) and stationary with autocovariances that decline at the hyperbolic rate k^{1-q} if $2 > q > 1$ (i.e. $0 < d < 1/2$). Notice that the hyperbolic decay of the memory parameter of an $FI(d)$ process, with $0 < d < 1/2$, contrasts with the exponential decay of the standard stationary ARMA of $I(0)$ process, enabling ARFIMA processes to model dependence between observations at long range. Likewise, in contrast with the $I(1)$ process, $FI(d)$ processes, with $0 < d < 1$, exhibit eventual mean reversions (see Cheung and Lai, 1993).

In the previous analysis we have assumed that the individual series were perfectly independent. However, some of the events to which individuals react are bound to be common though allowing for different individual responses. Other shocks, related to the individual well being, will be idiosyncratic. Denoting the common shock by W_t and the idiosyncratic shock by ε_t^i , the series will be partly dependent, being generated by:

$$y_t^i = c^i + \alpha^i y_{t-1} + \beta^i W_t + \varepsilon_t^i \quad (10)$$

where the series W_t and ε_t^i are assumed to be independent, ε_t^i is a white noise, and the α 's and the β 's are drawn from independent distributions. Then, y_t is $FI(d)$ where d is the largest of two terms (see Granger, 1980, section 3): $1 - q/2$ (from the ε component) and $1 - q - d_w$ (from the W component), where d_w is the degree of integration of W_t . However, to the extent that W_t represents 'news', conditional on an information set at $t-1$, $d_w = 0$ and therefore, $d = 1 - q/2$. Thus, the previous analysis remains valid and the null hypothesis that y_t follows a pure fractional process, i.e. an ARFIMA(0, d , 0), seems a plausible one. Indeed, as will be seen below, that specification is the preferred one for all of the series examined in this paper. Hence, q can be identified directly from the estimated value of d . Once q is estimated, it is possible to obtain the value of p that most nearly reconciles a pure fractional aggregate process with the theoretical autocovariances represented in (7). Therefore, by estimating the parameters of the $Beta(p, q)$ distribution, this procedure will allow us to obtain the probability mass of electoral support for each political party in the sample subject to analysis.

A final, yet crucial, issue to be addressed before moving onto the following sections, is how to construct a proxy variable for the aggregate partisanship, Y_t . For that, we follow BDP in using the log-odds of the average propensities to vote for a given party ($\bar{X}_t = \Sigma_i X_t^i / N$) where, by the law of large numbers, the probability limit of \bar{X}_t behaves in the limit as the average of the expectations of the unobservable binary variable X_t^i which takes a value of 1 if the i th voter supports a given party in the poll and 0 otherwise. Note that the log-odds of the average propensity, defined as $Y_t = L(\bar{X}_t) = \log(\bar{X}_t / (1 - \bar{X}_t))$ has support over the whole real line (approaching $\pm\infty$ as \bar{X}_t tends to 1 or 0), and is 0 when $\bar{X}_t = 0.5$, i.e. when the average propensity reflects the indifference of the electorate between supporting a given party or not. In this way, Y_t mimics the properties of the partisanship indicator defined by FJ. At this stage, it should be emphasized that our approach differs from the microfoundations provided by BDP where the variable capturing individual partisanship is directly defined as the log-odds of the expectation of the unobserved binary variable X_t^i , i.e. $L(X_{it}) = \log(X_{it} / (1 - X_{it}))$. As pointed out by one of the referees, the previous assumption is not strictly correct since the probability limits of the average log-odds, $N^{-1} \Sigma_i L(X_{it})$, and the log-odds of the average propensity, $L(\bar{X}_t)$, do not coincide. Insofar as our approach directly relies upon characterizing Y_t as the log-odds of the average propensity, leaving unspecified y_t^i , we avoid that problem. In effect, we are implicitly assuming that y_t^i is measured in such a way that its aggregated value across individuals gives rise to $L(\bar{X}_t)$ rather than imposing the log-odds transformation for X_t^i , as BDP do. However, to get a feeling about the extent to which the previous incorrect assumption by BDP can be taken as a sufficiently good local approximation to characterize the long memory properties, we carried out a small Monte Carlo experiment in order to check whether long-memory in \bar{X}_t is translated into long-memory in $L(\bar{X}_t)$. For that, we have simulated 1000 AR(1) processes with the AR parameters being drawn from a $Beta(p, q)$ distribution with $p = 0.9$ and $q = 0.6$. According to (7)–(9), it can be shown that the choice of those parameters replicates the persistence properties of an ARFIMA(0, 0.7, 0) process which, as will be discussed below, turns out to be a representative specification for our data. The sample size is $T = 600$. Then, for each of those AR(1) processes, y_t^i , we have used the logistic transformation $e^y / (1 + e^y)$, to map their original domain onto (0,1), as BDP assume for X_t^i . Next, we computed the average of the individual log-odds of the transformed series and the log-odds of the averages, and estimated d for both series. The number of replications for this comparison is 500. We found the mean-estimated value of d for the average of the individual log-odds was 0.701, in agreement with the choice of $q = 0.6$, whereas the estimate of d obtained from $L(\bar{X}_t)$ is 0.442. From this experiment, we conclude that, in spite of the fact that both estimates do not coincide, there is evidence that long memory is present in both functions of X_t^i . We also found that this property remains for other values of p and q . Thus, despite the presence of an incorrect step in the microfoundations of BDP 'model of political popularity, their claim that the log-odds of the average propensities follows a long-memory process seems to hold.

3. ESTIMATION AND TESTING PROCEDURES FOR FRACTIONALLY INTEGRATED PROCESSES

To test whether the aggregate time series of political support are well represented by a $FI(d)$ process, we make use of the Fractional Dickey–Fuller test (FD-F test) proposed by DGM (2002) which allows us to test the null of $I(1)$ against the alternative hypothesis of $FI(d)$, with $0 \leq d < 1$. This test turns out to be easily applicable in the time-series domain and fares better in finite

samples, in terms of power and size, than many of the other tests available in the literature. It can be interpreted as a Wald test in the time domain which provides a natural framework for estimating the memory parameter d in a rather simple way.

The FD-F test is based on the following unbalanced regression model which generalizes the one used in the standard Dickey–Fuller (D-F) approach by encompassing the hypothesis that a series x_t is either $I(1)$ or $FI(d)$ with $0 \leq d < 1$,

$$\Delta x_t = \rho \Delta^d x_{t-1} + u_t \quad (11)$$

where $\rho \leq 0$ and u_t is a linear covariance stationary process which, for simplicity, will be taken to be white noise in what follows. As in the D–F approach, deterministic terms, like a constant or a linear trend, can be appended to the previous regression. It is easy to prove that for $-2^{1-d} < \rho < 0$, x_t is a $FI(d)$ process with $0 \leq d < 1$, since the L.H.S. of (11) can be rewritten as $(\Delta^{1-d} - \rho L)\Delta^d x_t$ (with L denoting the lag operator) where the polynomial $\Pi(z) = (1 - z)^{1-d} - \rho$ has all its roots outside the unit circle. Thus, under the alternative hypothesis of $\rho < 0$, it follows that $\Delta^d x_t = e_t$, where $e_t = \Pi(L)^{-1} u_t$. On the other hand, under the null of $\rho = 0$, it becomes an $I(1)$ process; indeed a random walk. Therefore the null hypothesis of $I(1)$ corresponds to the case $\rho = 0$ versus the alternative of $FI(d)$, $0 \leq d < 1$, when $\rho < 0$. Being a Wald test, the intuition for the FD-F test is that, under the alternative hypothesis the appropriate filter of Δx_t will be $(1 - L)^{d-1}$, namely $\Delta x_t = \Delta^{1-d} e_t = e_t - (1 - d)e_{t-1} + (1/2!)(1 - d)(-d)e_{t-2} - \dots = \rho \Delta^d x_{t-1} + a_t$, so that the coefficient on $\Delta^d x_{t-1}$ in (11) can be interpreted as $(d - 1)$, i.e. the distance between the values taken by d under H_0 and H_1 , respectively. Further, if e_t happens to be i.i.d, it will be uncorrelated with $\Delta^d x_{t-1} (= e_{t-1})$, so that OLS yields a consistent estimator of ρ . On the other hand, if u_t happens to be a more general linearly autocorrelated process, DGM prove that (11) can be rewritten as:

$$\Delta x_t = \rho \Delta^d x_{t-1} + \zeta_p(L)\Delta x_{t-1} + \varepsilon_t \quad (12)$$

where $\zeta_p(L)$ is a lag polynomial of order p , where $p = o(T^{1/3})$ and ε_t is white noise. As in the earlier case, $\rho = 0$ and $\rho < 0$ correspond to the $I(1)$ null and the $FI(d)$ alternative, respectively. That version of the test, in line with its counterpart in the D–F setup, is denoted as the Augmented FD-F (AFD-F) test. As in the D–F approach for $I(1)$ against $I(0)$ processes, a simple test of $H_0 : \rho = 0$ can be obtained from either the normalized OLS estimator of ρ in (12) or from its associated t -ratio. In the sequel, we will concentrate on the properties of the latter test statistic while details on the properties pertaining to the normalized coefficient test can be found in DGM. Hence, application of OLS to (12) yields the following asymptotic distribution for the t -ratio of $\hat{\rho}_{ols}$ (denoted in short as t) under the null hypothesis:

$$t \xrightarrow{w} \frac{\int B_{1-d}(r)dB(r)}{\left(\int_0^1 B_{1-d}(r)dr\right)^{1/2}} \quad \text{if } 0 \leq d < 1/2,$$

$$t \xrightarrow{w} N(0, 1) \quad \text{if } 1/2 \leq d < 1$$

where $B(r)$ and $B_{1-d}(r)$ are standard Brownian and standard fractional Brownian motion, respectively, and \xrightarrow{w} denotes weak convergence. If a constant or a linear trend is included in

the model, the Brownian motions will correspond to their demeaned or detrended versions. Note that what matters in determining the asymptotic behaviour of the tests is the distance between the null and the alternative hypotheses. When this distance is large ($0 \leq d < 0.5$), the distribution is non-standard while when it is small ($0.5 \leq d < 1$), it is standard, i.e. asymptotically Gaussian. Also note that when $d = 0$, the asymptotic distribution derived by Dickey and Fuller (1979, 1981) is obtained. Critical values for the non-standard case for various sample sizes and a wide range of values of d have been tabulated by DGM (2002, Table I), allowing for the presence of a constant/trend in the model.

Obviously, it is often the case in practice that an estimated value of d would be needed in order to implement the test. In principle, any \sqrt{T} -consistent estimator of $d \in D \equiv [0, 1)$ could be used in regressions (11) or (12) to make the FD-F test feasible. Among them, the most popular ones are the maximum-likelihood estimator (MLE) due to Sowell (1992) and the minimum distance estimator (MDE) due to Tieslau *et al.* (1996). However, those estimators present shortcomings either due to their computational complexity or to just being defined for some subsets of D . In order to overcome those difficulties, we prefer to use the MDE of d proposed by Mayoral (2000) for the case where u_t is an ARMA (r, s) process, i.e. $u_t = \Phi_r(L)^{-1}\Theta_s(L)\varepsilon_t$, where $\Phi_r(L)$ and $\Theta_s(L)$ are lag polynomials with all roots outside the unit circle. The process is assumed to have an unknown mean (μ_0) after being differenced an integer number (m_0) of times. Hence, x_t follows the ARFIMA (r, d_0, s) process

$$\Phi_r(L)\Delta^{\varphi_0}(\Delta^{m_0}x_t - \mu_0) = \Theta_s(L)\varepsilon_t \quad (13)$$

such that $\varphi_0 \in (-3/4, 1/2)$, $d_0 = \varphi_0 + m_0$ and m_0 is the integer part of $(d_0 + 1/2)$. Thus, $d_0 \in (-3/4, \infty)$, i.e. a set that contains D . For example, if $d_0 = 0.7$, then $m_0 = 1$ and $\varphi_0 = -0.3$, whereas if $d_0 = 0.3$, then $m_0 = 0$ and $\varphi_0 = 0.3$.

Let $\Psi = (\phi_r, \theta_1, \dots, \theta_s)'$ be the vector containing the AR and MA parameters, $\lambda = (d, \Psi)'$ and $\zeta_j(\lambda)$, $j = 0 \dots \infty$, be the coefficients associated with the expansion $\Phi_r(L)\Theta_s(L)^{-1}\Delta^{d_0}$. Then, given the observations x_1, \dots, x_T , we can define the mean-adjusted residuals:

$$e_t(\lambda) = \sum_{j=0}^{t-m-1} \zeta_j(\lambda)[z_{t-j}(m) - \bar{z}(m)] \quad (14)$$

where $z_t(m) = \Delta^m x_t$ and $\bar{z}(m) = (1/T - m)^{-1} \sum_{t=m+1}^T z_t(m)$. Notice that $e_t(\lambda_0) = \varepsilon_t$. Also, since $z_t(m_0)$ is stationary and ergodic, the sample mean $\bar{z}(m_0)$ is a consistent estimator of μ_0 .

Next, define $V_{Te}(\lambda) = \sum_{i=1}^k \hat{\rho}_{ie}^2(\lambda)$ where $\hat{\rho}_{ie}(\lambda)$ is the sample i th-order autocorrelation associated with the residuals $e_t(\lambda)$, where $k = o_p(T)$. In practice, $k = T^{1/4}$ is a good choice. Then, the MDE of λ is defined as:

$$\hat{\lambda} = \arg \min_{\lambda \in \Lambda} V_{Te}(\lambda) \quad (15)$$

Notice that, since $e_t(\lambda_0) = \varepsilon_t$, the population autocorrelations are zero and $V_{Te}(\lambda)$ has a unique minimum. Moreover, it can be proved that $\hat{\lambda}$ turns out to be a \sqrt{T} -consistent, asymptotically normally distributed and efficient estimator of λ which has several advantages over other available estimators in the literature. First, when compared to estimators based upon the minimization of a weighted distance between the estimated and theoretical autocorrelations of the original series, x_t , as in Tieslau *et al.* (1996), it has the advantage of not being restricted to the stationary range of d , $-1 < d < 1/2$, for which only the autocorrelations of x_t exist. This is so since the proposed

MDE is based on the autocorrelations of the residuals which always exist. Secondly, relative to and the MLE proposed by Sowell (1992), it has the same asymptotic variance but is computationally much simpler and d is not restricted to lie in the stationary range.

Mayoral (2000) presents Monte Carlo evidence showing that the proposed MDE provides very accurate estimates of d even for $T = 100$, a sample size similar to the one used in this paper. For example, having simulated 500 replications of an ARFIMA (1,1,0) process with $\phi_1 = 0.6$ and $d = 0.4(0.8)$ the average estimated value of d is 0.401(0.789) and that of ϕ_1 is 0.57(0.58) for that sample size. Moreover it has two desirable properties. First, being \hat{d} a \sqrt{T} -consistent estimator of $d_0 \in (-0.5, 1.5)$, an interval that includes D , the t -ratio of $\hat{\rho}$ in (11) and (12) will behave asymptotically as a standardized normal variate. This is so since \sqrt{T} -consistency implies that the distance between \hat{d} and d_0 will be small, as happened to be the case when a hypothesized value of d , with $0 < 1 - d < 0.5$, was being used in implementing the FD-F test. Thus, when d is pre-estimated with the MDE, the asymptotic distribution is always a standardized normal variate. Second, the power of the test based on (12), for $d = \hat{d}$, is larger than the power obtained using other popular tests for fractional integration available in the literature, such as the non-parametric test developed by Geweke and Porter-Hudak (1983) or the frequency-domain one by Robinson (1994). For instance, for $T = 100$ and $d = 0.7$, the power of the FD-F test (with an estimated d) is 86%, against the 24% and 73% obtained for Geweke and Porter-Hudak's test and Robinson's test, respectively.

4. RESULTS

We examine the time series of percentage support of the five main Spanish political parties which have continued being active over the whole sample period. The data has been obtained from de Centro de Investigaciones Sociológicas (CIS) and it represents the answer to the following question: *Which party would you vote for if there were to be a Parliamentary Election tomorrow?*, which is posed once every quarter to a stratified sample of about 2500 individuals. Table I lists the names of the parties, their acronym, mean support (MS), the coefficients of variation (CV) and the max–min range over the sample period which runs from 1978(3) to 1999(4). Thus there are 87 quarterly observations, a sample size which is smaller than those used in other studies in this line of research. For example, the time series analysed by BDP (2000) for several OCDE countries have lengths ranging from 150 to 480 monthly observations. Before discussing the results, however, it should be noted that the total average support for the five considered parties only adds up to 48.7% of the total. The reason for this seemingly low coverage is that, on average, around 40% of the interviewed persons either did not declare support any given party or directly intended to abstain from voting. Moreover, before 1982, the party with the largest support was UCD, a centre party

Table I. Political parties and mean support

Party (acronym)	MS(%)	CV	Max–min (%)
Partido Socialista Obrero Español (PSOE)	26.44	0.41	13.2–45.4
Alianza Popular/Partido Popular (AP-PP)	13.44	0.53	3.1–26.4
Partido Comunista Español/Izquierda Unida (PCE-IU)	5.36	0.46	2.0–12.3
Partido Nacionalista Vasco (PNV)	1.05	0.21	0.4–2.2
Convergencia y Unión (CIU)	2.41	0.24	0.8–5.1

led by Prime Minister Suárez which won the first two general elections with an overall support close to 20%, and which suffered a severe defeat in 1982, to the point of causing its disappearance. Since the sample size for UCD only covers the 1978–1982 period (20 quarterly observations), it was excluded from the study.

It is worth noticing that the standard deviations of the voting intentions for the two small parties (CIU and PNV), which only present candidates for general elections at their regional constituencies (Catalonia and the Basque Country, respectively), are around one-fifth of their mean support and the range is fairly narrow. That indicates (see below) that a large proportion of their electorate belongs to the ‘militant’ class. For the three large parties (AP-PP, PSOE, PCE-IU) the share of ‘non-militants’ is larger, ranging from 7 percentage points for PCE-IU to about 20 percentage points for PSOE. Accordingly, our interpretation below of the degree of persistence in the poll time series for all parties would correspond to the fractions of their voters who react to unexpected changes in the information set. This fraction, in absolute terms, is of course much larger for the three large parties than for the other two smaller parties.

Figure 1(a) and (b) shows the raw data and the logistic transform of the series in deviations from its sample means. It should be noticed that an intercept was the only deterministic term which turned out to be significant when trying to fit an election cycle. Thus, the estimated models correspond to the specification in (13). Other variables such as the inflation and unemployment rates, and the length of the term in power, were also tried as controls but were never statistically significant. This result is in agreement with those reported by BDP (2000) who do not find an election cycle, except for the UK, in their study of partisanship data in eight major OCDE economies. Given that our sample only covers six general elections and three parties in power (UCD, from 1978 to 1982, PSOE from 1982 to 1996, and AP-PP from 1996 to the present), it seems natural that it will be difficult to capture those effects.

Table II reports standard tests of the hypotheses that the series are $I(0)$ and $I(1)$. We present the results for the Augmented Dickey–Fuller (ADF) and the Phillips and Perron (P–P) tests, for which the null hypothesis is $I(1)$, both with and without a deterministic trend, and also for the KPSS test which considers $I(0)$ as null hypothesis. If the series were $FI(d)$, with $0 \leq d < 1$, then both null hypotheses should be rejected, although it should be borne in mind that unit root tests have low power against non-stationary $FI(d)$ alternatives, particularly when sample sizes are not too large as it happens in our case. An asterisk indicates rejection at a 5% significance level. In general, it is found that both tests reject their respective null hypotheses, indicating that the presence of long-memory is quite embedded in the data.

Table III reports the values of the BIC information criterion for six alternative ARFIMA models estimated with the MD procedure described above. In all cases the preferred model was an ARFIMA $(0, d, 0)$, a result which agrees with interpretation of the disturbances in (10) as innovations capturing ‘news’.

Table IV, in turn, presents the result of applying the FD-F test (regression (11) with a constant term) to the poll data, together with the estimated value of d (second column) and the Box–Ljung $Q(k)$ statistic for k th-order autocorrelation in the residuals of the fitted model (third column), which are depicted in Figure 1(c). Note that in all cases the null of unit root is rejected against an $FI(d)$ with $0 < d < 1$. Moreover, the null hypothesis of uncorrelated residuals cannot be rejected, giving further support to the pure fractional noise process as an appropriate representation of the series at hand. For comparison, we also report the maximum likelihood (ML) estimator (\hat{d}_{ML}) of d implemented in Sowell (1992) (fourth column). This estimator is only valid when

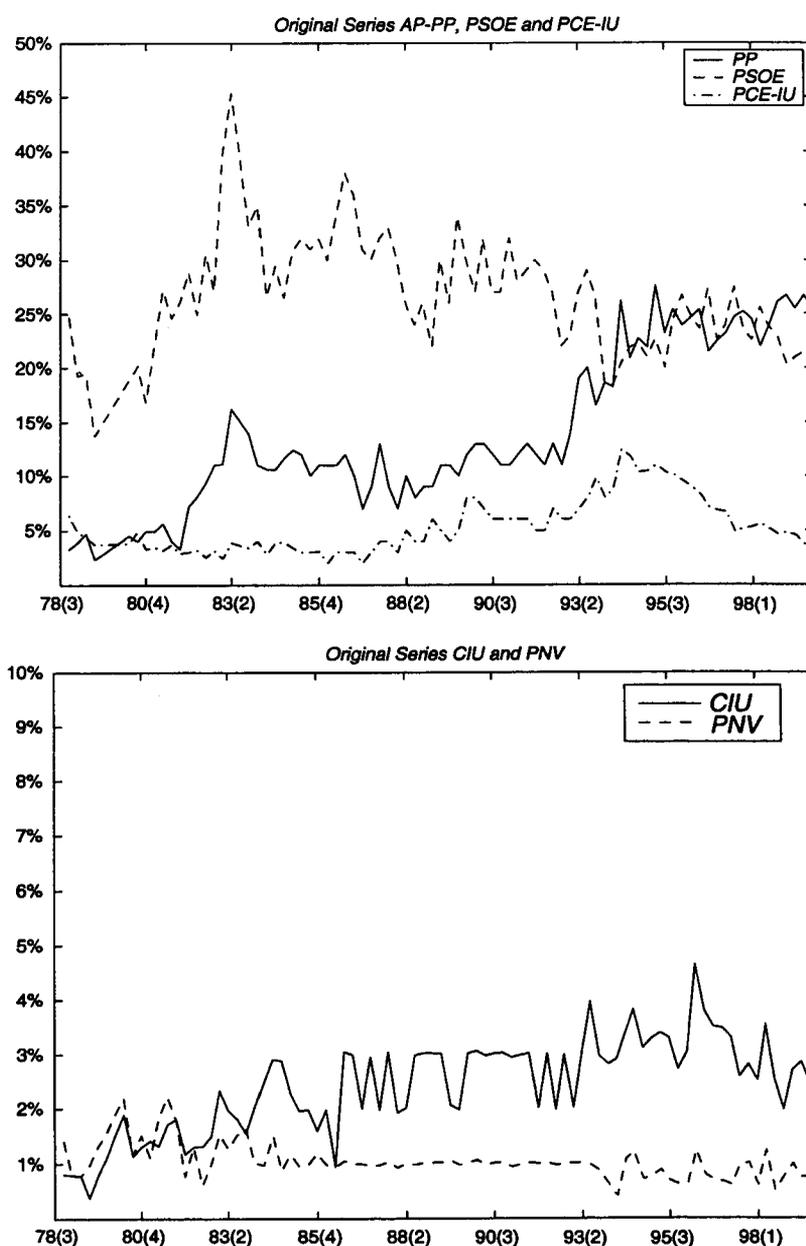


Figure 1. (a) Original series. (b) Log-odds transformation. (c) Fractionally differenced series

$-0.5 < d < 0.5$ and hence the reported estimates were computed by adding 1 to the estimate of d obtained with the first-differenced data. The results obtained with both methods are strikingly similar, a fact which offers further support to our use of the computationally simpler MD estimation method.

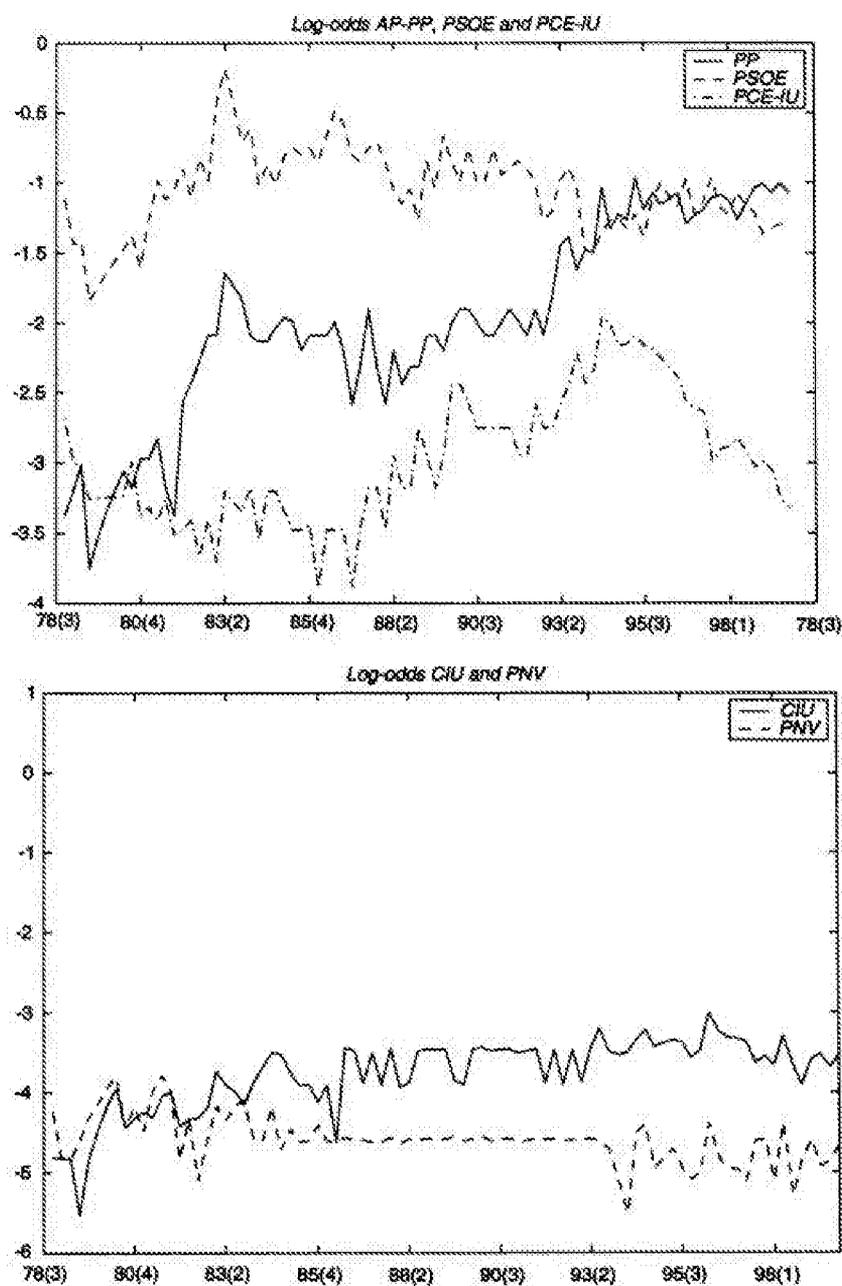


Figure 1. (Continued)

From the results in Table IV, it follows that in all cases, except for the PNV, the estimates of d are in the range $0.5 < d < 1$, giving rise to non-stationary, yet mean-reverting, $FI(d)$ processes. The highest value of d is found for AP-PP whilst the lowest value pertains to the

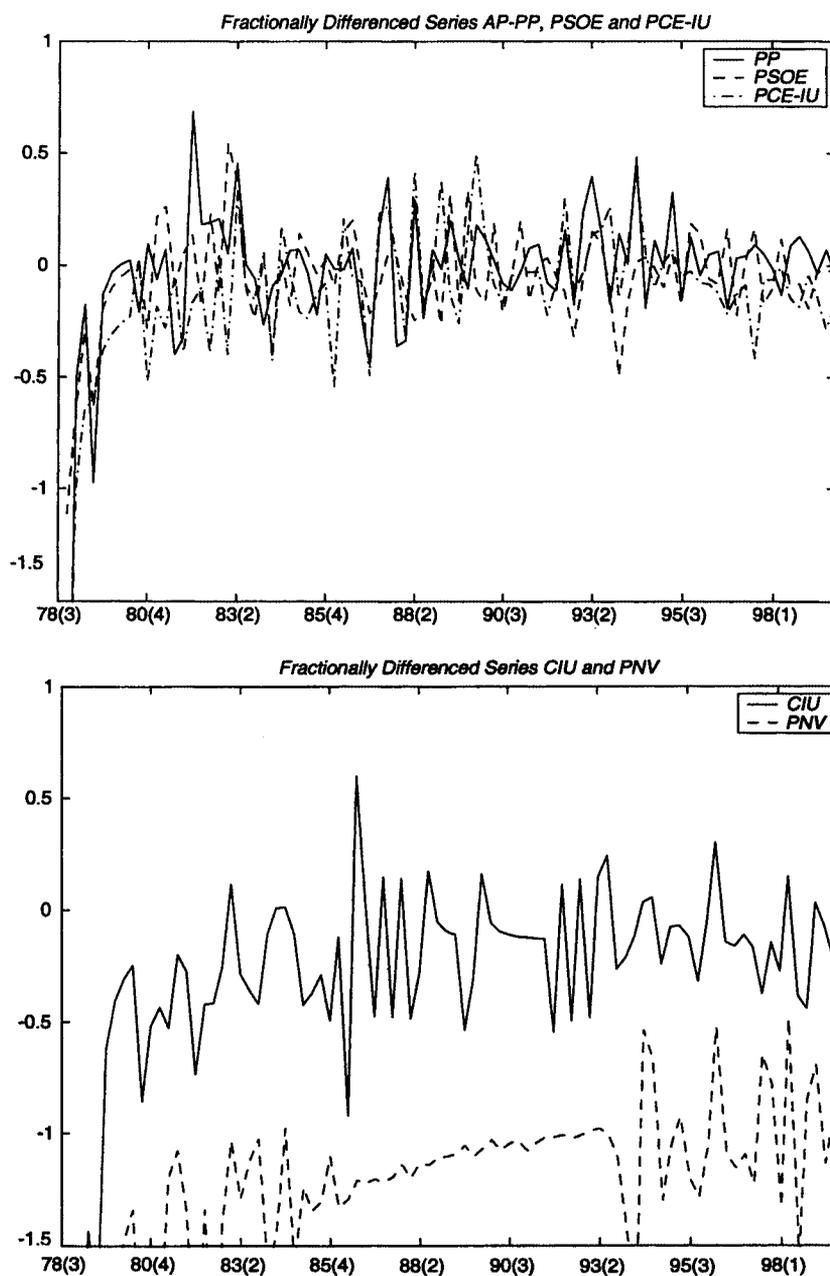


Figure 1. (Continued)

PNV. Except for the latter case, the results do not contradict the findings by BDP (2000, Table 3) who obtain estimates of d around 0.7 for the partisanship aggregate-level data in eight major OECD economies.

Table II. Stationary and unit root tests

	P-P $_{\mu}$	P-P $_{\tau}$	ADF $_{\mu}$	ADF $_{\tau}$	KPSS $_{\tau}$
AP-PP	-1.59	-3.67*	-1.80	-3.65*	4.67*
PSOE	-3.28*	-4.05*	-3.25*	-3.98*	4.53*
PCE-IU	-1.59	-3.43	-1.49	-2.63	5.78*
CIU	-3.79*	-6.80*	-2.87	-4.09*	3.22*
PNV	-4.29*	-7.05*	-3.98*	-4.19*	3.02*

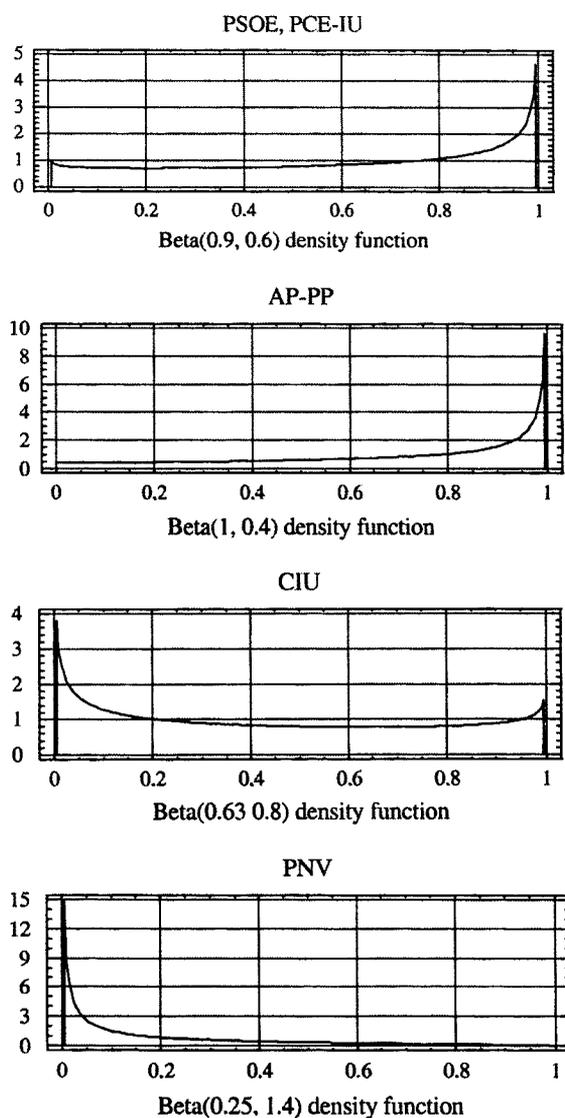
Table III. BIC values

	(0,d,0)	(1,d,0)	(2,d,0)	(1,d,1)	(0,d,1)	(0,d,2)
PSOE	-277	-269	-267	-264	-269	-264
AP-PP	-234	-224	-219	-219	-226	-217
PCE-IU	-292	-272	-230	-210	-273	-252
CIU	-298	-276	-270	-265	-276	-280
PNV	-980	-964	-953	-962	-964	-949

Table IV. FD-F test

	t	\hat{d}	Q(9)	\hat{d}_{ML}
AP-PP	-2.43*	0.83	3.56	0.81
PSOE	-2.67*	0.69	6.82	0.71
PCE-IU	-3.09*	0.74	12.78	0.72
CIU	-4.90*	0.59	8.67	0.60
PNV	-3.84*	0.33	10.73	0.31

As BDP (1997, p. 487) claim 'it seems that political commitment across the voting population might be a relatively stable sociometric constant that does not depend on factors such as left-right political orientation, etc.'. Thus, despite Spain's younger democracy and that the value $d \approx 0.7$ just represents an empirical regularity with no theoretical underpinning, our previous results for the three major Spanish parties clearly support the conventional findings in the literature. However, our finding that the long-memory parameter for the smaller regional (nationalistic) parties is lower than for the mainstream parties is an interesting result, since it may imply the existence of some significant differences in the behaviour of both types of supporter. Next, following the discussion in Section 3, we report those values of p which most nearly reconcile the *Beta* (p, q) distribution with uncorrelated shocks. We distinguish four cases corresponding to the rough estimates $d = 0.8$ (for AP-PP), $d = 0.7$ (for PSOE and PCE-IU), $d = 0.6$ (for CIU) and $d = 0.3$ (for PNV). For each case, the values of p that yield uncorrelated shocks, i.e. a pure fractional model, turn out to be 1.0, 0.9, 0.63 and 0.25 respectively. Figure 2 plots the shapes of the densities in each case. Assuming, for illustrative purposes, that a value of α above or equal to 0.8 corresponds to a committed voter, we can compute the probability masses associated to both types of voter from the above densities. In this way, we find that the proportion of committed voters is 50% for the political party (AP-PP) with the highest value of d , while it is 35% for those of an intermediate value (PSOE, PCE-IU), and 20% and 5% for those with the lowest values (CIU and PNV, respectively).

Figure 2. *Beta* (p, q) density function

Lastly, we have used the ARFIMA $(0, d, 0)$ models estimated above for prediction purposes. Multistep model forecasts for the period 1999(3)–2000(4) are shown in Table V. They have been computed using Ox version 2.10 (see Doornik, 1998) and the ARFIMA package version 0.95 (Doornik and Ooms, 1998) based on results by Beran (1994). The mean forecast errors for the period 1999(1)–1999(4), are 1.0%, –1.0%, 0.1%, 0.0%, –0.1% and 0.0% for AP-PP, PSOE, PCE-IU, CIU, and PNV, respectively.

Using the previous forecasts (adjusted by the mean forecast error for the previous election held in 1996) an admittedly bold exercise was undertaken to predict the number of seats in parliament

Table V. Forecasts

	AP-PP	PSOE	PCE-IU	CIU	PNV
99*(3)	25.23	21.50	4.69	2.70	0.83
99*(4)	24.98	21.78	4.79	2.67	0.86
00(1)	25.29	21.69	3.75	2.64	0.88
00(2)	25.07	21.86	3.89	2.62	0.89
00(3)	24.87	22.00	3.99	2.60	0.90
00(4)	23.07	22.11	4.06	2.59	0.90

in the last general elections called for 12 March 2000, at the time when the first draft of this paper was written, restricting the analysis to the three major parties. The number of seats in the 1996 general election was 156, 140 and 21 for AP-PP, PSOE, and PCE-IU, respectively, with the total number of seats in the Spanish parliaments being 350. Their mean support during the quarter preceding that election was 24.1%, 23.0% and 10.3% respectively, whereas the forecasts for 2000(1) were 25.3, 21.7 and 3.8, as shown in Table V. Thus, given the forecast for 00(1) and its associated standard error, we used a simple linear interpolation method to map mean support onto parliamentary seats. For example, since the mean forecast error for AP-PP was 1%, the point forecast for the number of seats of that party would be 170 ($= 26.3 * 156 / 24.1$). A confidence interval for the point forecast can also be constructed along similar lines mapping the standard error of the forecast on the political support onto seats. Using that approach for the other two parties yields the following distribution of seats: between 165 and 175 for AP-PP, between 123 and 129 for PSOE and between 6 and 10 for PCE-IU. Since overall majority is achieved with 176 seats, our predictions therefore were that: (i) AP-PP would win the elections being close to reaching overall majority, and (ii) a potential coalition between PSOE and PCE-IU would not be enough to reach victory in the elections.

The results of the general elections turned out to be 183, 125 and 8 seats for AP-PP, PSOE and PCE-IU, respectively. Thus, except for AP-PP, which got overall majority by a margin of 7 seats, the remaining outcomes almost coincided with our point estimates and were well within the confidence intervals. Hence, fractionally integrated processes seemingly provide a powerful tool in predicting election outcomes. Nonetheless, the absence of a completely proportional rule in the Spanish electoral system, where D'Hont's rule applies, ought to put some dose of caution on the use of the above simple methodology for forecasting the outcomes of general elections. A deeper analysis of the specific characteristics of the electoral system at hand, in order to improve forecasting, will be the subject of further research.

5. CONCLUSIONS

The results in this paper support the view that ARFIMA processes with a value of the fractional differencing parameter, d , between 0.6 and 0.8 provide an appropriate characterization of persistence in the time series of opinion polls for mainstream political parties in Spain. This result is in broad agreement with the findings for opinion polls for major parties in countries with longer democratic traditions than Spain. The fact that d is fairly similar across countries seemingly supports the idea that such a distribution might represent a stable constant of mass political behavior in democratic societies. By assuming a certain distribution of the attributes of committed and pragmatism in a voting population, we are able to interpret d as a measure of the relative size of the

'committed' and 'floating' populations of supporters for each party. In particular, we find that the current party in power (PP) has the largest proportion of 'committed' voters, around 50%, whereas the main opposition party (PSOE) only has 35%. It should be noted, however, that those shares are larger than the ones obtained for the smaller regional parties (CIU and PNV) whose estimated ds are between 0.3 and 0.6, giving rise to shares of committed voters of between 5% and 20%. Our interpretation of the latter result is that those small shares only pertain to the 'non-militant' fraction of their supporters, being their 'militant' populations much more important than for the mainstream parties.

Finally, the estimated ARFIMA models perform well in forecasting the results of the last general elections in Spain, held in March 2000, using information on the opinion polls up to the previous quarter.

ACKNOWLEDGEMENTS

We are very grateful to James Davidson, Miguel Delgado, a co-editor, two anonymous referees and seminar audiences at the Universidad de Alicante, CEMFI (Madrid), EUI (Florence), ESEM 2001 (Lausanne) and the University of Montreal for useful comments and suggestions. Financial support from DGESIC Grant SEC01-0890 is also gratefully acknowledged. The usual disclaimer applies.

REFERENCES

- Achen CH. 1975. Mass political attitudes and the Survey response. *American Political Science Review* **69**: 1218–1231.
- Beran J. 1994. *Statistics for Long Memory Processes*. Chapman & Hall: New York.
- Box-Steffensmeir JM, Smith RM. 1996. The dynamics of aggregate of partisanship. *American Political Science Review* **90**: (September): 567–580.
- Byers D, Davidson J, Peel DA. 1997. Modelling political popularity: An analysis of long-range dependence in opinion polls series. *Journal of the Royal Statistical Society Series A* **160**: 471–490.
- Byers D, Davidson J, Peel DA. 2000. The dynamics of aggregate political popularity: Evidence from eight countries. *Electoral Studies* **19**: 49–62.
- Cheung Y, Lai K. 1993. Long memory in foreign exchange rates. *Journal of Business and Economic Statistics* **11**: 93–101.
- Chrystal KA, Peel D. 1987. What can economics learn from political science and vice versa. *Economic Review* **76**(2): 62–65.
- Dickey DA, Fuller WA. 1979. Distribution of estimators of autoregressive time series with a unit root. *Journal of the American Statistical Association* **74**: 427–431.
- Dickey DA, Fuller WA. 1981. Likelihood ratio tests for autoregressive time series with a unit root. *Econometrica* **49**: 1057–1072.
- Dolado JJ, Gonzalo J, Mayoral L. 2002. A fractional Dickey–Fuller test for unit roots. *Econometrica* **70**: 1963–2006.
- Doornik JA. 1998. *Object-oriented Matrix Programming using Ox 2.0*. Timber-lake Consultants Ltd.: London.
- Doornik JA, Ooms M. 1998. A package for estimating, forecasting and simulating Arfima models. www.nuff.ox.ac.uk/Users/Doornik: Oxford.
- Eisinga R, Franses PH, Ooms M. 1999. Forecasting long memory left-right political orientations. *International Journal of Forecasting* **15**: 185–199.
- Franklin CH, Jackson JE. 1983. The dynamics of party identification. *The American Political Science Review* **77**(4): 957–973.
- Geweke J, Porter-Hudak S. 1983. The estimation and application of long memory time series models. *Journal of Time Series Analysis* **4**: 221–238.

- Granger CWJ. 1980. Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* **14**: 227–238.
- Granger CWJ, Joyeux R. 1980. An introduction to long-range time series models and fractional integration. *Journal of Time Series Analysis* **1**: 15–30.
- Holden K, Peel D. 1985. An alternative approach to explaining political popularity. *Electoral Studies* **4**: 231–239.
- Lippi M, Zaffaroni P. 2000. Contemporaneous aggregation of linear dynamic models in large economies. Preprint. Università La Sapienza, Rome.
- Mayoral L. 2000. A new minimum-distance estimator of ARFIMA processes. Preprint, Universidad Carlos III. Revised version (2002) available upon request.
- Robinson PM. 1994. Efficient tests of nonstationary hypotheses. *Journal of the American Statistical Association* **89**: 1420–1437.
- Sowell FB. 1992. Maximum likelihood estimation of stationary univariate fractionally-integrated time-series models. *Journal of Econometrics* **53**: 165–188.
- Tieslau M, Schmidt P, Baillie R. 1996. A minimum-distance estimator for long-memory errors. *Journal of Econometrics* **71**: 249–274.