Testing for Fractional Cointegration: the Relationship between Government Popularity and Economic Performance in the UK

James Davidson*
Cardiff University

July 2000

Abstract

This paper investigates the relationship between the quarterly opinion poll lead of UK governments over the period 1955-1996, and a set of economic indicators. The hypothesis of a causal link between these variables often debated, but there is a difficulty in testing the link by conventional econometric methods. These require either stationarity or the I(1) property, but there is strong evidence from a number of different studies that opinion poll series are fractionally integrated, being nonstationary but also mean-reverting.

This paper tests the hypothesis of fractional cointegration using bootstrap methods. It first discusses the problem of defining a cointegrating relationship between series that may not have the same order of integration, and suggests a generalized cointegration model that might account for this case. The bootstrap tests of the regular and generalized cointegration hypotheses make use of a size correction that compensates for the biases due to estimating the parameters of the model for the bootstrap replications.

The tests reveal no evidence of a link between the political and economic cycles, a conclusion that reinforces the results of earlier work suggesting that the political cycle is generated by the internal dynamics of the opinion formation process. The findings are reinforced by a Monte Carlo study, showing that the methods have ample power to detect cointegrating relations, even in cases where the residuals are noisy and persistent

1 Introduction

A substantial literature has accumulated over recent decades, seeking both theoretical and econometric links between economic conditions and the popularity of democratic governments. Leading contributions are Goodhart and Bhansali (1970), Nordhaus (1975), Frey and Schneider (1978), Pissarides (1980), Minford and Peel (1982), Holden and Peel (1985), Rogoff and Sibert (1989). The evidence from econometric studies, treating this as a conventional time series modelling problem, has been at best equivocal. For example, Pissarides (1980) uses the time series techniques suggested by Davidson et. al. (1978) and finds some nominally significant correlation between government popularity and economic indicators (growth, inflation, unemployment, the exchange rate and tax rate). However, his equation does not have much predictive power. While plenty of anecdotal evidence can be cited in support of either view, whether government popularity follows the economic cycle remains an unresolved question.

More recent research has found that for a wide range of countries and democratic political systems, party support is a fractionally integrated process. See for example Byers, Davidson and Peel (1997, 2000), Box-Steffensmeier and Smith (1996), and Dolado, Gonzalo and Mayoral (2000).

*Research supported by the ESRC under award L138251025. Email: davidsonje@cf.ac.uk

Byers et. al. (1997), henceforth referred to as BDP, show that for the UK that the monthly Gallup series for Conservative and Labour support can be well modelled as ARFIMA(0,d,0) with d around 0.75. In other words, the series is covariance nonstationary, but also not a random walk, tending to return from excursions away from the median.[1] In their paper, BDP propose a model to account for these findings based on the aggregation of heterogeneous poll responses, appealing to a well-known result of Granger (1980). The model accounts for the magnitude and duration of swings in aggregate opinion as due to the particular mix of committed and floating voters in the population. The innovations in the process are assumed to be news, of both the economic and non-economic variety. The BDP model therefore accounts for the cyclical behaviour of opinion by the internal dynamics of the aggregate opinion-formation process.

This explanation contradicts the view that swings in support follow economic indicators over the cycle. BDP suggest instead that the impact of economic variables be tested by examining the correlations between the innovations in the support series (the fractional differences) with innovations in economic indicators, and perform a test of this type. They find, at best, very slight effects of this kind; see BDP's Table 7. BDP explain this finding by noting that opinion polls aggregate the heterogeneous opinions of voters who perceive economic circumstances differently. Borrowers and depositors take a different view of the interest and inflation rates, for example. The so-called 'North-South divide' in the UK, and the contrasting fortunes of manufacturing and service industries, show how unemployment may fail to move the employed majority however distressing it may be for the unemployed themselves.

However, a formal test of the relationship still remains wanting. Of the two statistical approaches to testing for time series relationships in common use, the correlation approach and the cointegration approach, neither is valid when the data in question are fractionally integrated. Since the support series is nonstationary, ordinary tests of significance are subject to the well-known 'spurious correlation' critique. On the other hand, cointegration analysis relies on tabulations of the distribution of certain functionals of Brownian motion, and accordingly are based on the assumption that the time series have variances diverging at the rate n. In the case of a fractionally integrated or I(d) process, this rate is n^{1-2d}, and the limit processes are not Brownian motion but fractional Brownian motion. The Brownian functionals defining the limit distributions depend on d, and the usual cointegration tests are inappropriate. In fact, it is not possible to generate asymptotically pivotal tests unless d is known.

The present paper gives a test of the hypothesis using the bootstrap to overcome the problems with conventional tests. The theory of the tests reported here is discussed at length in Davidson (2000). Section 2 of the paper presents the data set to be analysed. Section 3 considers some issues in the modelling of relationships in such data. Sections 4 and 5 describe the bootstrap test procedure, and Section 6 interprets the results. Section 7 gives the results of an extended power evaluation, and Section 8 concludes.

2 The data set

The data for the present study are quarterly observations for the period 1955:2 to 1996:4. They are shown in raw form in Figure 1, and in Figure 2 as deviations from linear trend, fitted by least squares. The party support data are taken from the monthly Gallup poll series, and 'Lead' is measured as the difference between Conservative and Labour percentage support in periods

[1] BDP model the series for $\log[\bar{X}_t/(1-\bar{X}_t)]$ where $\bar{X}_t$ is the sample average support. This process is defined on $(-\infty,+\infty)$ and a random walk is a logically feasible representation. In practice the range of variation of the $\bar{X}_t$ series is such that the logistic transformation is nearly linear, and the same model explains either series equally well.

Dependent Variable: Lead	Sample: 1955:2 1996:4		
	Coefficient	Std. Error	t-Statistic
Inflation	0.093	0.155	0.600
Real Earnings	0.074	0.372	0.200
Real GDP	0.268	0.575	0.467
TB Rate	1.578	0.318	4.948
Unemployment	2.867	0.686	4.179
Constant	-21.56	18.79	-1.147
Trend	-0.432	0.220	-1.960
R-squared	0.342	Adjusted R-squared	0.317
Durbin-Watson	0.575	F-statistic	13.89
Residual ADF	-4.512	F-statistic excl. Trend	13.54

Table 1: Regression of Lead on Economic Indicators

of Conservative government, and the difference between Labour and Conservative in periods of Labour government.

Five indicators have been chosen as possible economic explanations of support. A linear trend is added to the data set on the grounds, both theoretical and from inspection of the series, that it is the trend deviations that would be likely to influence support for better or worse. The results of running the 'cointegrating' regression appear in Table 1 and Figure 3.

The null hypothesis to be tested is, in effect, the BDP hypothesis. The only relation between Lead and economic variables, according to this hypothesis, is that the innovations of the processes may be correlated. As noted, the available evidence gives grounds to doubt even this connection. The alternative hypothesis is that the stochastic trend in Lead is driven by, and hence cointegrated with, the trends in the measured economic indicators. There is a third hypothesis, that Lead does not follow the BDP model but is driven by other, unmeasured, economic variables. This possibility must remain unresolved by the present analysis, though it remains open to test by the methods of this paper if the variables can be identified.

Since there could be a temptation to treat this as a conventional cointegration analysis, it is worth noting the outcome in this case. The residual ADF statistic in Table 1 (computed with 4 lags) is not far from the 10% critical value, according to MacKinnon's (1991) tables. While the hypothesis of non-cointegration could not be rejected in the conventional I(1) framework, the margin would be slender enough to leave room for doubt. One may guess that the p-value for this test would be no more than 12-15% which, taken at face value, might be construed as weak evidence for the alternative.

However, both the I(0) and I(1) assumptions are contradicted by the results of the univariate ARFIMA modelling exercises reported in Table 2. This table shows ARFIMA(p, d, q) models for each series in the data set, chosen to maximise the Schwarz selection criterion, subject to the side condition that residual autocorrelation is insignificant by the Box-Pierce Q test for 12 lags. Each of these estimates was computed by differencing the data to satisfy the stationarity/invertibility condition $|d| < 0.5$, and then adding 1 to the estimate of d so obtained. The second Box-Pierce test provides evidence of possible ARCH-type nonlinear dependence (McLeod and Li 1983), which of course the ARFIMA framework cannot account for. However, these models are generally adequate and parsimonious. The Lead variable, in particular, is well represented by the ARFIMA$(0, d, 0)$ model with d significantly exceeding 0.5, indicating the series to be nonstationary, but also significantly less than unity. The conventional cointegration test is therefore also in doubt.

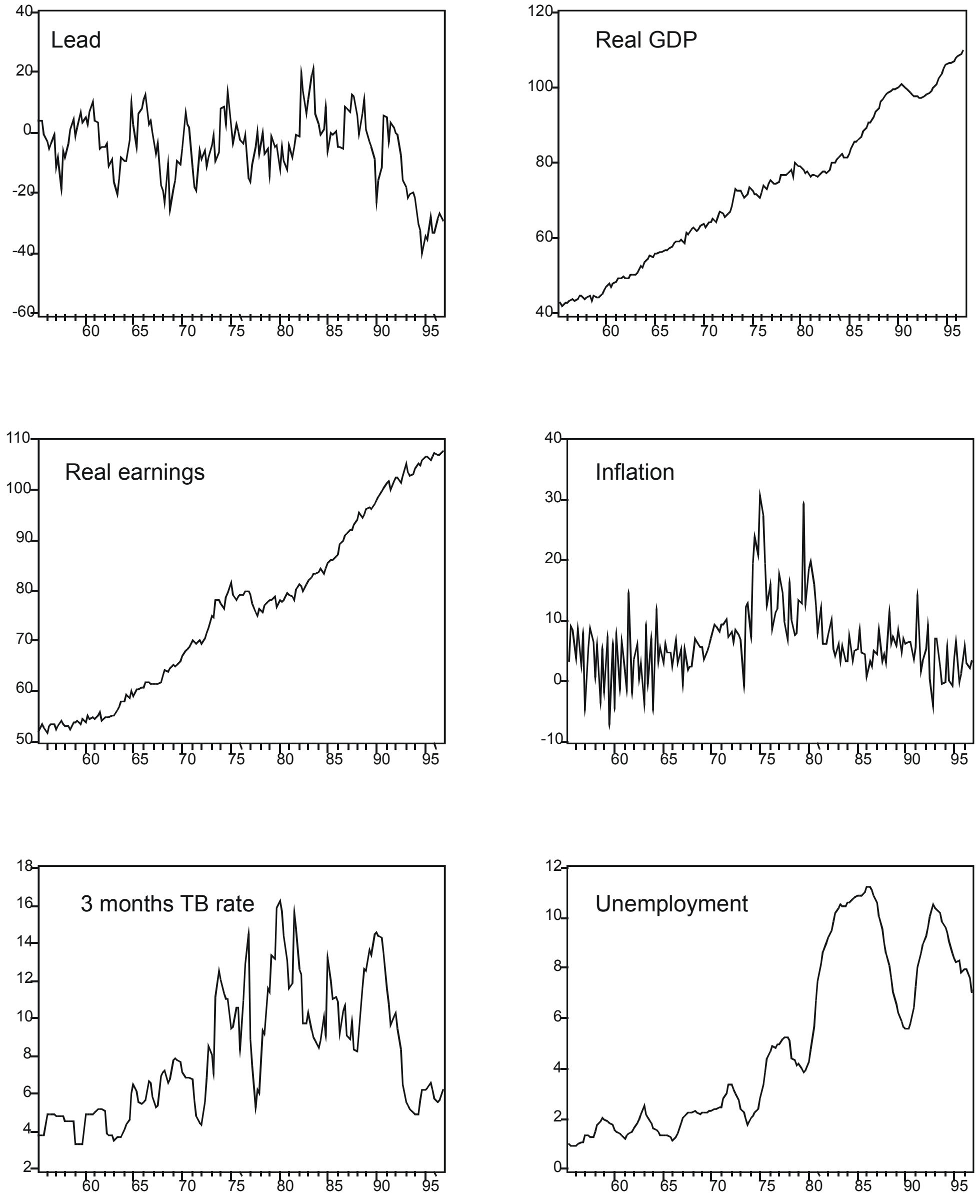

Figure 1: Data

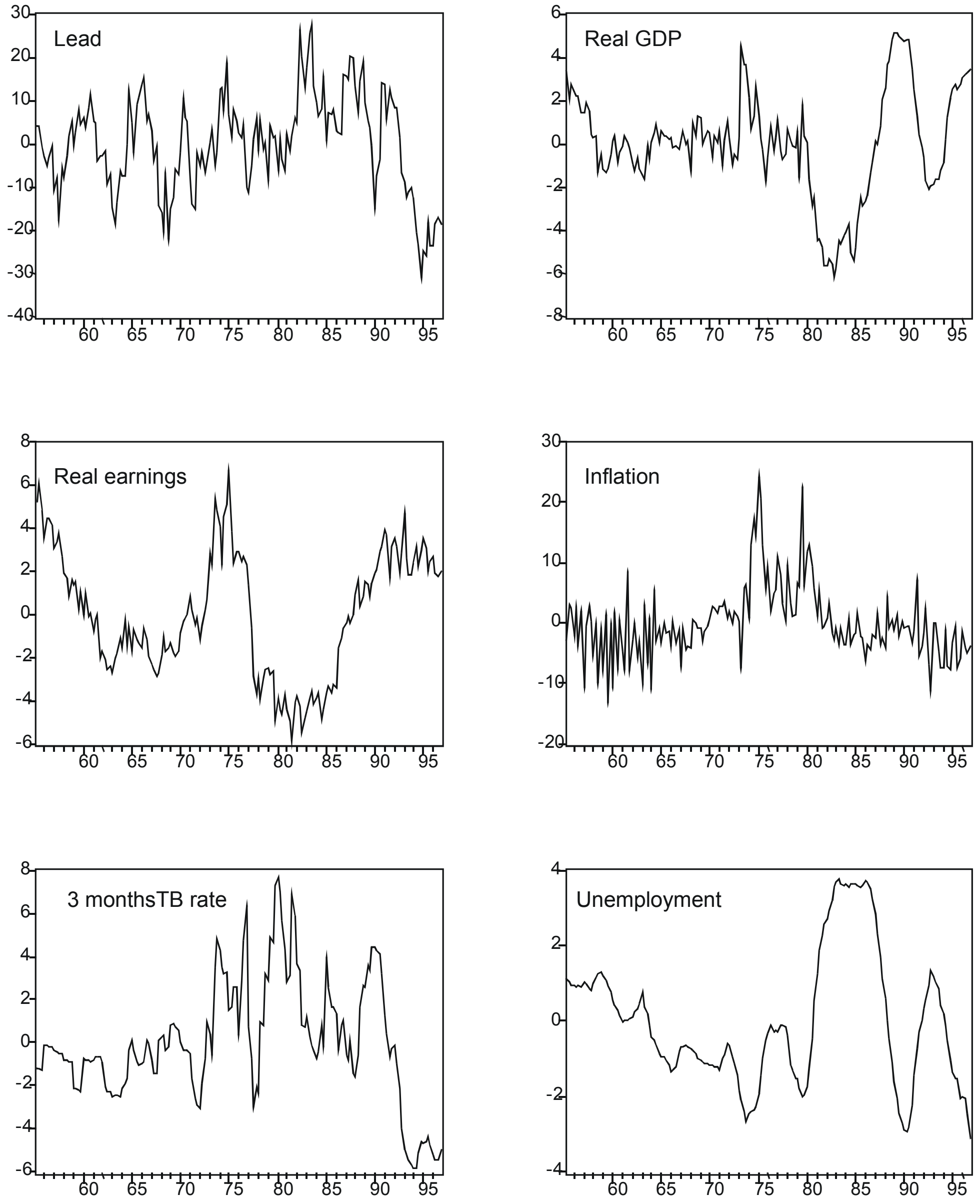

Figure 2: Data as deviations from trend

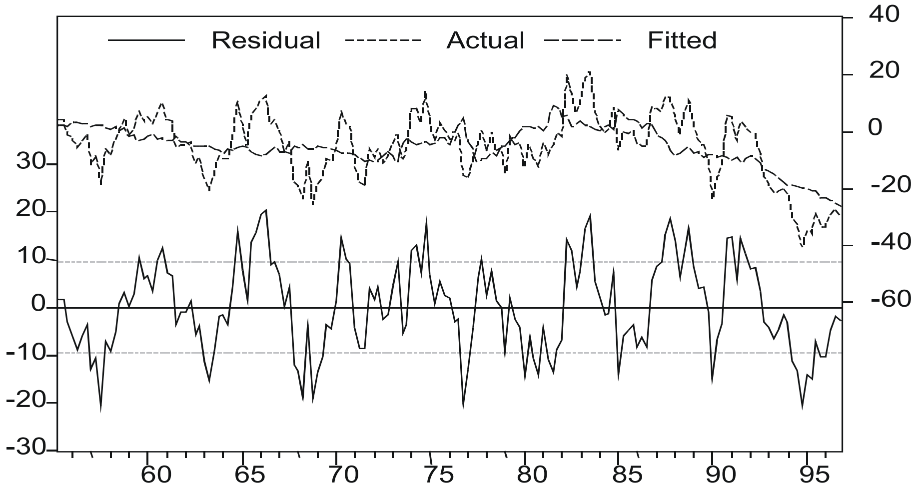

Figure 3: Regression of Lead on Indicators

	Lead	Unempl	Infl.	TBR	GDP	RE
d	0.765	1.169	0.664	0.626	0.978	0.920
	(0.066)	(0.150)	(0.092)	(0.107)	(0.060)	(0.079)
p	0	2	1	2	0	0
q	0	0	0	1	0	0
ARMA Coefficients:	-	0.518	0.473	0.397	-	-
		(0.16)	(0.135)	(0.105)		
	-	0.208	-	0.466	-	-
		(0.107)		(0.094)		
	-	-	-	0.814	-	-
				(0.038)		
Constant	0.084	-0.22	5.942	4.924	38.98	46.59
Trend	-0.061	0.062	0.007	0.037	0.401	0.353
$Q(12)$ - levels	12.09	17.76	19.90	12.32	17.58	12.92
$Q(12)$ - squares	11.40	19.80	17.61	9.20	9.85	12.71

Table 2: Best ARFIMA(p,d,q) models of the data set (std. errors in parentheses)

3 Models of Fractional Cointegration

From the point of view of establishing a relationship, the results of Table 2 present some unexpected problems. While unemployment, the interest rate and earnings all have estimated d insignificantly different from unity, this is not true of either lead or the interest or inflation rates. These are significantly mean-reverting, although nonstationary ($1/2 < d < 1$). Is it possible that variables with different orders of integration can be cointegrated?

To answer this question, consider the fractional vector ECM model given in Davidson (2000). This takes the form

$$\left[\boldsymbol{B}(L) + \boldsymbol{\alpha}\boldsymbol{\beta}'(\boldsymbol{K}(L)^{-1} - I)\right]\boldsymbol{\Delta}(L)\boldsymbol{x}_t = \boldsymbol{D}(L)\boldsymbol{\varepsilon}_t. \tag{3.1}$$

where

$$\boldsymbol{\Delta}(L) = \text{diag}\{(1-L)^{d_1}, \dots, (1-L)^{d_N}\} \tag{3.2}$$
$$\boldsymbol{K}(L) = \text{diag}\{(1-L)^{b_1}, \dots, (1-L)^{b_N}\} \tag{3.3}$$

where $d_1, \dots, d_N$ are any nonnegative reals (assume $d_1 \geq \cdots \geq d_N$ without loss of generality), $0 \leq b_i \leq d_i$, and $\boldsymbol{B}(L)$ and $\boldsymbol{D}(L)$ are $N \times N$ polynomial matrices whose characteristic roots are strictly outside the unit circle. In the usual way, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are $N \times r$ matrices with rank r. This system generates N series integrated to orders $d_1, \dots, d_N$, such that

$$\boldsymbol{\Delta}(L)\boldsymbol{x}_t = \boldsymbol{w}_t \sim \text{I}(0) \tag{3.4}$$

(defining $\boldsymbol{w}_t$). If $\boldsymbol{\alpha} = \boldsymbol{\beta} = \boldsymbol{0}$ these are noncointegrated, but if $r > 0$ it is required, to balance the equation, that

$$\boldsymbol{\beta}'\boldsymbol{K}(L)^{-1}\boldsymbol{w}_t \sim \text{I}(0). \tag{3.5}$$

If $b_i > 0$ for one or more i, this implies cointegration. This set-up encompasses a wide range of possible models. If $b_i = b$ and $d_i = d$ for all i it corresponds to the system proposed in Granger (1986), and if $b = d = 1$ then it reduces to the Johansen (1988, 1991) style VECM. More generally, we can pick out a number of other cases yielding a possible modelling framework.

The first of these is where $d_i - b_i = a \geq 0$ for each i, which implies that

$$\boldsymbol{\beta}'\boldsymbol{x}_t \sim \text{I}(a). \tag{3.6}$$

If $a > 0$, this is the case often called fractional cointegration, in which the cointegrating residual is long memory and possibly even nonstationary, but has a lower order of integration than its constituent variables. It is clear that with $b_i > 0$, this model cannot have property (3.6) except subject to additional restrictions. As discussed in Davidson (2000), either $d_1 = d_2$ or the top row of $\boldsymbol{\beta}$ must be equal to $\boldsymbol{0}$, so that x_{1t} is not cointegrated with the other variables. It is possible that this set-up could describe the present case, since the data set contains three (plausibly) I(1) series. In other words, the trends in GDP, unemployment and real earnings cannot individually drive the trend in Lead, but a combination of these could, at least in principle, do so. We do not yet consider whether such a model would be behaviourally plausible, merely note the possibility.

A second case where model (3.1) could generate cointegated series is where $b_i = b \leq \min_{1 \leq i \leq N} d_i$ for all i, which, to ensure the equation balances, implies that

$$[(1-L)^{-b} - 1]\beta'\boldsymbol{w}_t \sim \text{I}(0). \tag{3.7}$$

Dependent Variable: Lead	Sample: 1955:2 1996:4		
	Coefficient	Std. Error	t-Statistic
Inflation	-0.117	0.154	-0.759
Real Earnings*	0.129	0.496	0.259
Real GDP*	0.162	0.900	0.179
TB Rate	1.566	0.328	4.948
Unemployment*	3.988	1.723	2.314
Constant	-19.11	30.42	-0.628
Trend	-0.472	0.395	-1.193

R-squared	0.263	Adjusted R-squared	0.235
Durbin-Watson	0.496	F-statistic	9.52
Residual ADF	$-.235$	F-statistic excl. Trend	8.63

Table 3: The generalized cointegration model

This model has the peculiarity that the cointegrated series are not the elements of $\boldsymbol{x}_t$ themselves, but the fractional differences of orders $d_i - b$.[2] This case will be referred to as *generalised cointegration*, to make the distinction with simple cointegration in which linear combinations of the measured variables have a lower order of integration, as in (3.6). This set-up allows imposes no restrictions on $\boldsymbol{\beta}$ to ensure cointegration. It allows cointegration to be defined between arbitrary sets of I(d) variables, and so resolves the main limitation of the fractional model as an econometric modelling device.

Again, whether this is economically and behaviourally plausible is a matter for consideration. There is nothing unusual in having the simple difference of a variable appear in an economic relationship. For example, the (log-) price level contains (at least) the same information as the level of inflation, but the latter variable is customarily assumed relevant to agents' decisions. While economic models do not normally assign the same role to fractional differences, this is simply because such a modelling strategy has never been entertained. There seems to be no inherent reason why they should not do so. Just as the price level is relevant to some decisions and its rate of change to others, in a representative-agent framework, so may the fractional difference of a trending variable contain the relevant information for a decision involving a particular planning horizon. In turn, this could be reflected in the degree of persistence of the target variable. The question of primary interest must be whether such relationships are discoverable in the data.

The result of running the regression on the present data after semi-differencing is shown in Table 3. The variables marked with a * have been semi-differenced to have a d of 0.765, based on the models in Table 2. The filtered (and also detrended) series are shown, with the originals for comparison, in Figure 4. On the conventional criteria, this regression is actually somewhat inferior to the original in Table 1, and offers little support for the generalized cointegration approach in the present context. It will nevertheless be of some interest to apply the test in this setting, one in which the cointegration hypothesis presents at least no logical difficulties.

4 Test Procedure

The bootstrap tests applied here are described in detail in Davidson (2000). The main feature of the procedure is to draw bootstrap replications of the model in (3.1) under H_0, such that $\boldsymbol{\alpha} = \boldsymbol{\beta} = \mathbf{0}$, and so generate the null distributions of two regression-based test statistics, the

[2] Note that the orders of integration of the cointegrated series are indeterminate unless we impose that the linear combination is I(0).

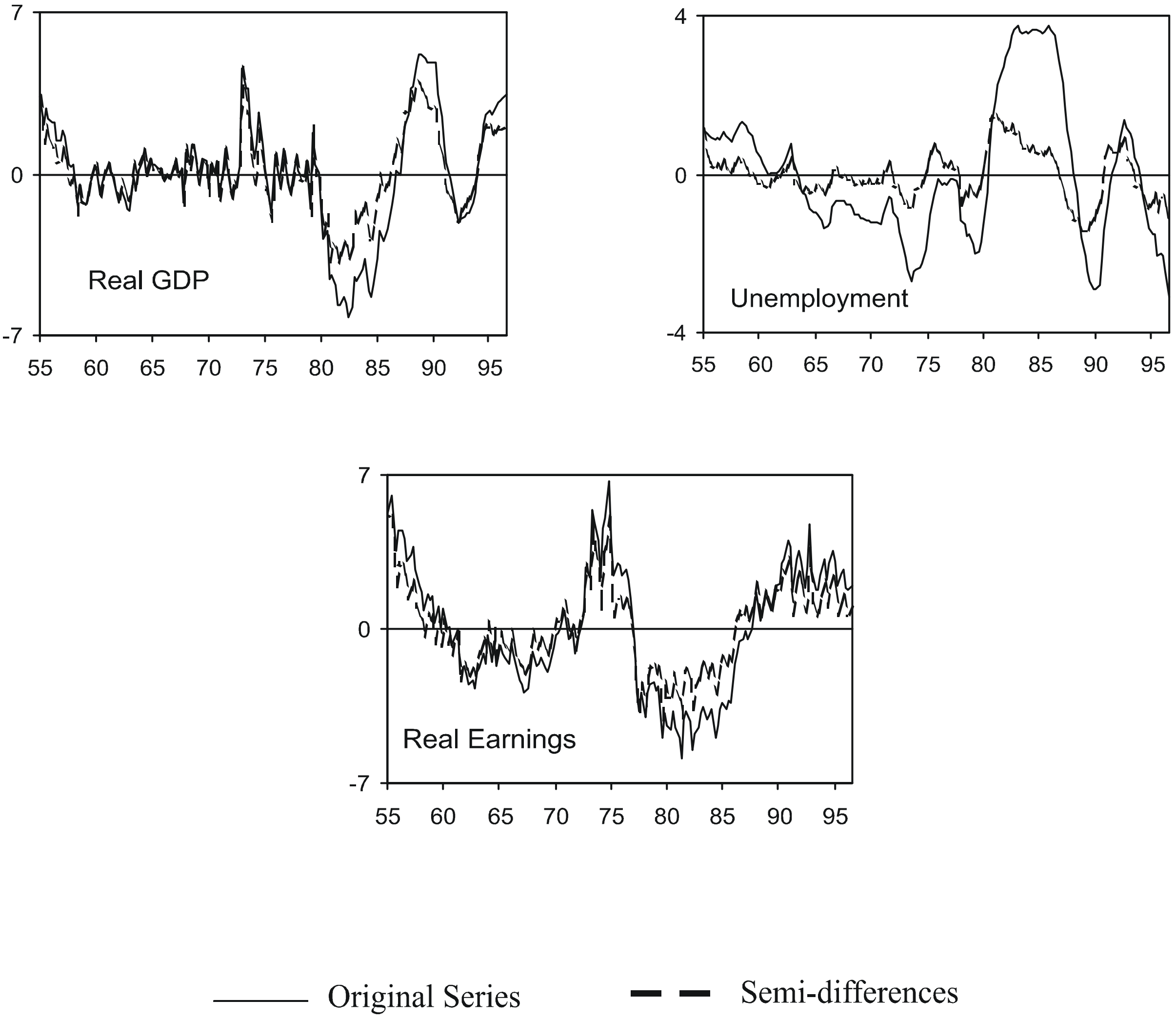

Figure 4: Semi-differenced series, $d = 0.765$

F statistic for goodness of fit (excluding trend) and the Durbin-Watson statistic. The actual statistics yielded by the regression in Table 1 are located in these empirical distributions to yield asymptotically valid p-values. The estimated value of d, from the first column of Table 2, is used to generate the I(d) series representing Lead, which is non-cointegrated with the regressors by construction, but whose increments reproduce the observed correlation structure with those of the regressors under H_0. The bootstrap draws are conditioned on the actual sample values of the regressors. This yields potentially more powerful tests than would bootstrapping the complete data set, noting that the conditional distributions must have smaller dispersion than the unconditional ones.

The test statistics employed are not asymptotically pivotal, meaning that they depend on nuisance parameters under H_0; specifically, the values of d and the autocovariances of the data increments. While there exist well known fixes to correct for these nuisance parameters in tests for conventional I(1)/I(0) cointegration, of which the 'augmentation' of the Dickey-Fuller statistic is the best-known, there are no such fixes that can generate statistics not depending on the d values. A bootstrapping approach is therefore unavoidable. In this framework, the dependence on the covariance parameters is allowed for, not by computing a modified statistic, but by generating the appropriate bootstrap distribution, by estimating the DGP of the increments of Lead under H_0.

Let $\boldsymbol{w}_t$ be the I(0) vector defined in (3.4). Because the regressors $\boldsymbol{x}_{2t} = (x_{2t}, \ldots, x_{Nt})'$ are to be held conditionally fixed, it is necessary to estimate a dynamic equation for w_{1t} containing both $w_{1,t-1}, \ldots$ and $\ldots, \boldsymbol{w}_{2,t+1}, \boldsymbol{w}_{2t}, \boldsymbol{w}_{2,t-1}, \ldots$, where the ellipses represent lags of total length to be specified. The inclusion of the leads as well as lags of the regressors is to allow for the fact that w_{1t} could Granger-cause $\boldsymbol{w}_{2t}$, which is not ruled out even if the regressors were weakly exogenous. With this structure, however, with lags suitably chosen, the residuals from the regression should be, asymptotically, both serially independent and totally independent of the regressors. Resampling from the empirical distribution of these residuals, and then passing them back through the same filter in reverse, should yield a bootstrap sample having the same correlation structure under H_0 as the original series, asymptotically, and the resulting test distributions should depend on the nuisance parameters in just the right way.

There is one caveat to be observed in this procedure. The test as described, in which the best-fitting dynamic equation is chosen by the usual consistent model selection criteria, should be correctly sized asymptotically, because if H_0 is true the correct model is chosen with probability 1 in the limit. However, such a test would have limited power, because when cointegration does exist, this long-run relation willl contaminate the short-run dynamics, and the best model must contain a large number of leads and lags. The problem is avoided by choosing a deliberately parsimonious model, with short leads and lags, which should capture the weak dependence under H_0 but avoid this contamination. In practice, there is a trade-off of advantages between size and power that can only be resolved in the light of experience and simulation studies.

Another feature of these tests is the opportunity to test different null hypotheses. Being regression based, these tests may appear directly comparable with the Engle-Granger or Phillips-Perron residual-based tests. This is true in the sense they can only test for cointegrating rank 0 against cointegrating rank > 0, but since they entail structural modelling of the short-run dynamics, they have much in common with system-based tests like Johansen's eigenvalue tests. If the null hypothesis is simulated by imposing independence between w_{1t} and $\boldsymbol{w}_{2t}$ — in effect, by excluding $\boldsymbol{w}_{2t}$ and its leads and lags from the dynamic regression — the test distributions generated are appropriate to the hypothesis that x_{1t} and $\boldsymbol{x}_{2t}$ are independent, not merely non-cointegrated. A similar test would be obtained by enforcing block-diagonality of the short-run component in the Johansen VECM.

We refer to the hypothesis of non-cointegration, with short-run correlation unrestricted, as the weak null, and the hypothesis of independence, with no relationship at all frequencies, as the strong null. The ability to test alternative dependence hypotheses promises to make these test procedures more informative than their conventional counterparts. This is one reason why they may be useful even when the latter are available, in the context of I(1) data.

5 Size Correction

The tests are consistent and asymptotically correctly sized, subject to the dynamic model under the null being correctly specified. In finite samples, however, there is evidence from simulations of fairly severe bias. To demonstrate the problem, let P_{n0} denote the probability measure associated with the distribution of a test statistic t when H_0 is true. If t_n is the realized value of the statistic, the bootstrap procedure estimates the p-value $g_n = P^*_{n0}(t \leq t_n)$, where P^*_{n0} is the distribution of t in the bootstrap replications. Note that $P^*_{n0} \neq P_{n0}$, because various nuisance parameters have been replaced by estimates in the bootstrap computation of t_n.

Defining the p-value corresponding to an exact test as

$$h_n = P_{n0}(t \leq t_n)$$

note that g_n and h_n are both random variables whose distribution is derived from that of t_n. However, for any choice of P_{n0} the relation

$$P_{n0}(h_n < x) = x \tag{5.1}$$

holds by construction, because t_n is a random drawing from P_{n0}. In other words, h_n is uniformly distributed on $[0, 1]$. Let H_n denote the c.d.f. of g_n, such that

$$P_{n0}(g_n \leq x) = H_n(x). \tag{5.2}$$

If the distributions are continuous, then $H_n : [0, 1] \mapsto [0, 1]$ is an increasing homeomorphism with $H_n(0) = 0$ and $H_n(1) = 1$, and from (5.2) it has the property

$$h_n = H_n(g_n). \tag{5.3}$$

Suppose the distribution of the data under H_0 were known, and random drawings could be taken from it. H_n could be computed, as accurately as desired, by the empirical c.d.f. of g_n in Monte Carlo replications of the whole bootstrap procedure, including the fitting of the ARFIMA model of y_t and the estimation of the short-run dynamics. In view of (5.1), by using h_n computed from (5.3) in place of g_n an *exact* α-level test could be constructed, given sufficient replications to estimate the distributions with arbitrary accuracy.

Of course, the distribution of the data is unknown, but since the DGP can be given parametric form and consistently estimated, Monte Carlo can also be used to construct a c.d.f. $\hat{H}_n$ (say) by taking drawings from this estimated distribution. The test based on $\hat{H}_n(g_n)$ is not exact and is still only an asymptotic approximation. However, it *would* be exact, in finite samples, in the case where the estimates of the nuisance parameters were identical with those of the true DGP generating P_{n0}.

We call this a size correction since, by construction, the test criterion 'reject if $\hat{H}_n(g_n) < \alpha$' must lead to rejection with probability α in Monte Carlo draws from the estimated DGP used to generate $\hat{H}_n$. This fact is useful in evaluating the power estimation carried out below in Section 7, since the DGPs under the alternative hypotheses reduce to the same null DGP by setting the parameters to their null values. However, the method is effectively equivalent to the prepivoting procedure of Beran (1988).[3] The statistic g_n can be thought of as an asymptotically pivotal statistic, given that its asymptotic distribution is $U(0,1)$ on H_0, regardless of the distribution of the underlying statistic. The statistic $\hat{H}_n(g_n)$ represents the proportion of bootstrap draws falling below g_n, and so is, in effect, the bootstrap p-value corresponding to g_n. Following Beran's analysis, the corrected test can therefore be said to have a true rejection probability on H_0 that differs from the nominal rejection probability by an error of $o(n^{-1/2})$.

6 Results

Using the fitted univariate ARFIMA models reported in Table 2 to provide estimates of the d parameters, bootstrap tests were perfomed of each of the four hypotheses, in other words, the strong and weak forms of regular and generalized cointegration. In the latter case, the three regressors with estimated d values exceeding 0.765 were partially differenced to match this value, as in Figure 4, while the other two were left in their original form. To estimate the short-run dynamics under H_0, the system dynamics are convoluted with the univariate time series models in Table 2 for the sake of parsimony, as described in Davidson (2000). Thus, to test the strong null hypothesis the ARFIMA residuals for Lead were resampled and back-filtered through the ARFIMA model to give the bootstrap samples. In the test of the weak null, the ARFIMA residuals were further modelled by regression on the lags of the regressors of orders -2 through $+2$, as well as 2 own-lags. This distribution was resampled, and passed back through the same filter in reverse.

The tests were performed with 5000 bootstrap replications, although note that this relatively large number does not imply a precisely estimated p-value. The test is asymptotic and the approximation depends on sample size, with $T = 167$ in this case. It does however render the sampling error small enough to ignore, so that the tests are directly comparable with conventional asymptotic tests.

To obtain the corrected p-values, the following procedure was replicated 5000 times. A sample drawing from the estimated distribution of Lead was obtained, from the distribution used to construct the bootstrap test under either the true null or the weak null hypotheses, as appropriate. Treating this as the 'true' data set, the bootstrap test was carried out using 500 replications. This test includes the steps of estimating the value of d for the series by maximum likelihood using the ARFIMA$(0, d, 0)$ specification, and fitting the short-run dynamics. The empirical c.d.f. of p-values so obtained represents the function $\hat{H}_n$ defined in the last section, and is used to read off the entries in the second of the two columns in Table 4. Figure 5 shows these estimates for each model, the broken line indicating the uniform c.d.f.

To perform the generalized cointegration test, the series for unemployment, real GDP and real earnings were semi-differenced, as described in Section 3. It should be emphasized here that the choice of common d value adopted here is arbitrary, to the extent that fractionally integrating Lead and the other variables 'up' to a larger common value (such as 1.16) is observationally equivalent. The only difference would be that the putative cointegrating residual would have longer memory than otherwise.[4]

It is evident that, on all these measures, there is not even slender evidence of a cointegrating relationship. The set of economic indicators chosen may be incomplete, and for example the

[3] The only difference is that in the present implementation the empirical c.d.f. $\hat{H}_n$ and the bootstrap p value g_n are computed in successive runs, instead of simultaneously. This is solely a matter of programming convenience, since the extra time needed to compute g_n is trivial.

[4] The test distributions would of course be different in the two cases. It is a reasonable conjecture that the test power should be invariant to this choice, but whether this is so would require further research to establish.

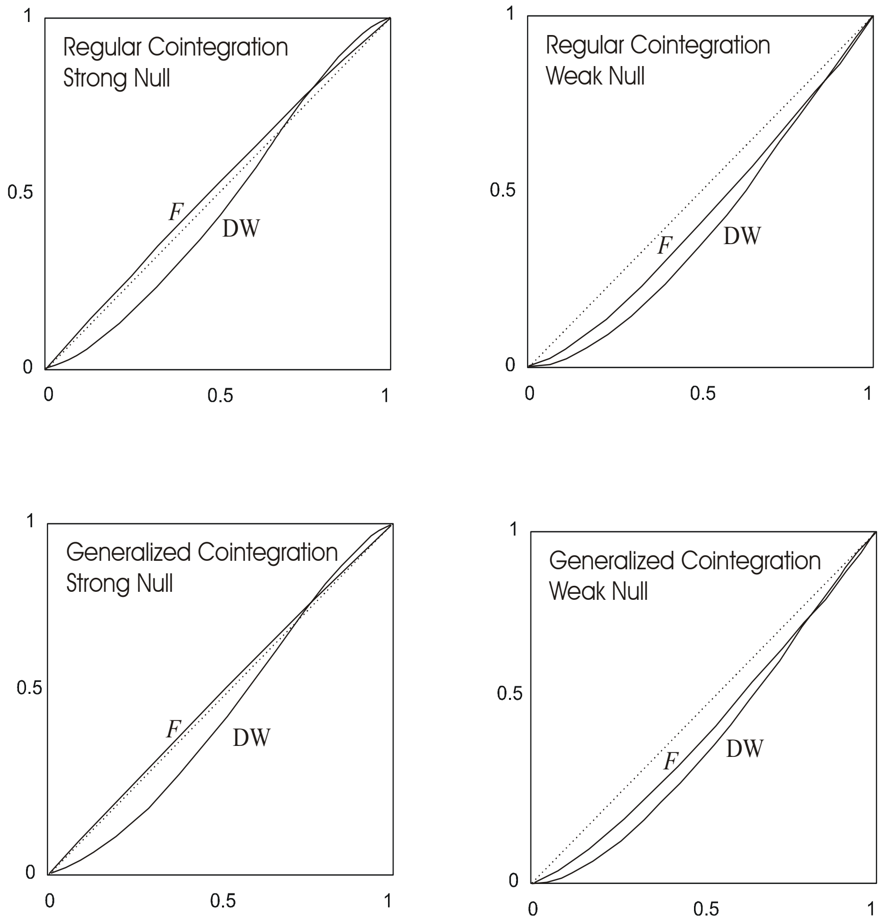

Figure 5: $\hat{H}_n$ functions, estimated from 5000 Monte Carlo replications.

		F		DW	
		raw	corrected	raw	corrected
Regular cointegration	Observed value	13.54		0.575	
	p-value, strong null	0.518	0.533	0.524	0.451
	p-value, weak null	0.635	0.544	0.687	0.572
Generalized cointegration	Observed value	16.04		0.64	
	p-value, strong null	0.492	0.511	0.297	0.208
	p-value, weak null	0.594	0.475	0.453	0.285

Table 4: Cointegration test results

tax rate and exchange rate indicators used by Pissarides (1980) have not been considered here. However, the variables included should on any basis be regarded as important. One would expect at least some mild evidence of a relationship, if in fact it existed. While alternative models are clearly open to test on the same lines, this evidence clearly favours either the dominance of purely non-economic factors, in explaining the trend, or an explanation on the lines proposed by BDP.

7 Power Evaluation

One of the virtues of the bootstrap approach is that a power evaluation can be undertaken relevant to the specific model under test. This is done by essentially the same procedure used to construct the size corrections of Section 5. In this case, however, artificial processes are constructed to represent a residual, to which is added a linear combination of the regressors to form a data set for which the null hypothesis is false. These models are chosen to be as close as possible to the null cases for which the size corrections were constructed, the same model of the fractional difference component being used in each case.

The experimental procedure is as follows:

1. As for the bootstrap procedure, the Lead variable is fractionally differenced according the model in Table 2. The differences are whitened and projected on the differenced regressors, as described in Section 4, to generate a shock series.

2. In each Monte Carlo replication, the shock series is randomly resampled. In a test of the weak null, it is back-filtered through the estimated VAR from step 1 to simulate the correlation structure. These steps are also identical to those performed in the bootstrap test.

3. Letting the series generated at step 2 be denoted $\{e_t\}$, the artificial dependent variable is constructed as

$$y_t(s,d) = s(1-L)^{-d}e_t + \boldsymbol{z}_t'\boldsymbol{b} \tag{7.1}$$

   where $\boldsymbol{b}$ is an arbitrary vector to form the 'explained' part of the model, and d and s are values to be varied experimentally. In practice, $\boldsymbol{b}$ was chosen as the coefficients from the regression on the observed y_t, from Table 1 or Table 3, as appropriate. d is set in the range from 0 up to 1, and

$$s = C\frac{SD(e_t)}{SD(\boldsymbol{w}_{2t}'\boldsymbol{b})}$$

   (recall that $\boldsymbol{w}_{2t}$ are the I(0) fractional differences of $\boldsymbol{z}_t$) where SD denotes the sample standard deviation and C a proportionality factor, to be varied.

4. The bootstrap test is performed on the artificial model, using the size corrections. The actual procedure is simulated to the point of estimating d for the dependent variable by maximum likelihood, although using the ARFIMA$(0,d,0)$ model rather than by a specification search.

Steps 2-4 were replicated 1000 times, and the power of the test estimated by the proportion of cases in which the estimated p-value fell below 0.05. There are four cases to consider, the strong and weak hypotheses, for the regular cointegration and generalized cointegration models, respectively, and for each of these cases, five values of d and four values of C were used. The

	$C=1$			$C=2$			$C=4$			$C=8$		
d	F	DW	R^2	F	DW	R^2	F	DW	R^2	F	DW	R^2
1	1	0.11	0.84	0.98	0.12	0.59	0.61	0.12	0.29	0.19	0.10	0.10
0.75	1	0.7	0.96	1	0.65	0.88	0.98	0.46	0.64	0.55	0.23	0.32
0.5	1	1	0.99	1	1	0.95	1	0.98	0.84	0.98	0.75	0.57
0.25	1	1	0.99	1	1	0.97	1	1	0.90	1	1	0.69
0	1	1	1	1	1	0.98	1	1	0.91	1	1	0.71

Table 5: Test powers: regular cointegration, strong null

	$C=1$			$C=2$			$C=4$			$C=8$		
d	F	DW	R^2	F	DW	R^2	F	DW	R^2	F	DW	R^2
1	0.80	0.44	0.97	0.42	0.29	0.59	0.15	0.19	0.27	0.08	0.15	0.08
0.75	1	0.95	0.96	0.94	0.84	0.88	0.43	0.50	0.65	0.14	0.19	0.33
0.5	1	1	0.99	1	1	0.96	0.97	0.99	0.84	0.50	0.61	0.58
0.25	1	1	0.99	1	1	0.97	1	1	0.90	0.95	0.96	0.69
0	1	1	0.99	1	1	0.98	1	1	0.91	1	1	0.72

Table 6: Test powers: regular cointegration, weak null

	$C=1$			$C=2$			$C=4$			$C=8$		
d	F	DW	R^2	F	DW	R^2	F	DW	R^2	F	DW	R^2
1	0.99	0.04	0.61	0.64	0.06	0.29	0.23	0.08	0.09	0.10	0.09	0.02
0.75	1	0.35	0.88	0.99	0.30	0.66	0.59	0.15	0.32	0.19	0.07	0.11
0.5	1	0.99	0.95	1	0.96	0.85	0.98	0.72	0.58	0.61	0.18	0.24
0.25	1	1	0.97	1	1	0.90	1	1	0.70	0.99	0.80	0.36
0	1	1	0.98	1	1	0.91	1	1	0.72	1	1	0.40

Table 7: Test powers: generalized cointegration, strong null

	$C=1$			$C=2$			$C=4$			$C=8$		
d	F	DW	R^2	F	DW	R^2	F	DW	R^2	F	DW	R^2
1	0.15	0.18	0.58	0.05	0.12	0.25	0.05	0.09	0.06	0.07	0.09	0.01
0.75	0.64	0.58	0.87	0.12	0.23	0.64	0.07	0.10	0.30	0.08	0.06	0.07
0.5	0.99	0.99	0.96	0.40	0.74	0.85	0.14	0.28	0.57	0.13	0.08	0.22
0.25	1	1	0.97	0.87	0.99	0.90	0.49	0.77	0.70	0.44	0.30	0.35
0	1	1	0.98	0.99	1	0.91	0.94	1	0.73	0.93	0.85	0.40

Table 8: Test powers: generalized cointegration, weak null

powers are shown in Tables 5–8, which also give the average values of the R^2 from the 1000 regressions, to give an idea how well these regressions are fitting in each case.

A number of considerations need to be borne in mind in reviewing these results. First, they are of course entirely specific to the data set and sample size in question, noting that $\boldsymbol{z}_t$ in (7.1) represents the observed data, on which all simulations are conditioned, just as for the tests themselves. The point is of course that the same evaluation can be performed, in principle, for any empirical problem.

Second, the strong and weak hypotheses have been set up in such a way that the nulls are exactly true in the cases $\boldsymbol{b} = \boldsymbol{0}$. Thus, for the weak hypothesis (Tables 6 and 8), the fractional differences of y_t in (7.1) are correlated with those of $\boldsymbol{z}_t$ whether or not $\boldsymbol{b} = \boldsymbol{0}$. In the strong null cases (Tables 5 and 7), they are independent when the null is true. Comparison of the two cases aims to show the cost in terms of test power of the need to take account of high-frequency correlation under the null.

Third, in the results in Tables 5 and 6, since the regressors are close to I(1) the same will be true of the artificial dependent variables. The alternatives investigated here might be thought of as corresponding to that state of the world in which the d for Lead has been mismeasured. Fourth, finally, note that the results in Tables 7 and 8 do not tell us anything about the generalized cointegration model, as such, but merely about the ability of the test to detect cointegration in fractionally integrated processes, where the common d is 0.765.

The main features of these results are the generally good power of both the tests to detect cointegration in this setting, and also the equal or superior power of the F test in nearly every case. The asymptotic analysis given in Davidson (2000) has a bearing on these findings. It is shown there that the DW-type test has power against alternatives with the residual $d < 1$. Remarkably, the F test can in some cases detect a relationship even when the disturbance is I(1), provided the shocks are small enough.

8 Conclusion

This paper has sought evidence for a connection between the popularity of governments and economic indicators over the business cycle, and has failed to find any. Since negative findings of any sort can leave readers in doubt about the quality of the evidence, it is as well to emphasize what conclusions can be drawn here.

First, note that the strong null hypothesis is that of independence. The statistics will in general will have their distributions shifted to the right in the presence of correlation between the increments of the processes, even in the absence of cointegration.[5] The test is not consistent against such alternatives, but such correlations should at least shift the median of the test distribution. In the present case, Table 4 shows that the observed statistics lie close to the medians of the strong-null distributions.

Second, while these may not be exact tests in finite samples, exact tests of level α can be constructed by rejecting only if the largest possible p-value, by choice of the unknown nuisance parameters, is less than α (see Dufour 1999). Clearly, no such test can reject the non-cointegration hypothesis, and we can therefore treat these results as exact at any chosen significance level. Even by admitting type 1 error probabilities as large as 20%, the hypothesis cannot be rejected.

Third, while the set of economic series chosen for the test may omit some important ones, those included are undeniably important. It has been shown that these tests have power against alternatives in which the residuals are long-memory or nonstationary — even I(1). The main

[5] These properties is demonstrated in Davidson (2000).

implication of these properties is that omission of important factors should not in general mask an existing relationship. If any economic trend factors are omitted, we are forced to the conclusion that they must be orthogonal to those included, and it is not at all obvious what these factors might be.

Fourth, in focusing on the formalities of cointegration testing we have not commented on the numerical magnitudes of the regression coefficients in Tables 1 and 3, but obviously these have dubious implications. Of the coefficients with large t values, the positive relationship between unemployment and popularity appears bizarre, although we can account for it anecdotally by pointing to, for example, the catastrophic collapse of the 1992-97 Conservative government's popularity, in step with recovery from recession. Historians of the period will explain this decline in terms of misbehaviour by politicians, internal divisions, and a loss of confidence following the exit from the ERM. We know that such intangible factors matter. What the present results show is that objective economic conditions have an insignificant role by comparison.

In summary, then, this study can be claimed to provide clear evidence in support of the BDP hypothesis, that local trends in popularity have quite different causes relating to the aggregation of sampled opinions. Economic events send different messages to different individual voters, and aggregating their reactions to them has unpredictable effects. Minor events are important if voters agree about them, major events may be appear to be ignored in the aggregate if voters disagree. Whatever the actual mechanism of opinion filtering, the effect is to scramble the original message so effectively that it is undetectable in statistical tests The message for governments may be that while the economy is undoubtedly important, the constituencies of winners and losers under any change of policy have to be offset against one another, and the effects are hard to disentangle.

References

Baillie, R.T. (1996) 'Long memory processes and fractional integration in econometrics' *Journal of Econometrics* 73, 5-59

Beran, R. (1988) 'Prepivoting test statistics: a bootstrap view of asymptotic refinements' *Journal of the American Statistical Association* 83, 687-697.

Box, G. E. P., and D. A. Pierce (1970). 'The distribution of residual autocorrelations in autoregressive-integrated moving average time series models' *Journal of the American Statistical Association* 65, 1509-26.

Box-Steffensmeier, J. M. and R. M. Smith, (1996) 'The Dynamics of Aggregate Partisanship", *American Political Science Review* 90 (September), 567-580

Byers, D., J. Davidson, and D. A. Peel, (1997) 'Modelling Political Popularity: An Analysis of Long-range Dependence in Opinion Poll Series', *Journal of the Royal Statistical Society* Series A, 160, 471-490

Davidson, J. (2000) 'A model of fractional cointegration, and tests for cointegration using the bootstrap'. Working Paper, at http://www.cf.ac.uk/carbs/davidsonje.

Davidson, J., Hendry, D.F., Srba, F. and Yeo, 5. (1978) 'Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom', *Economic Journal* 88, 661–92.

Dufour, J.-M. (1999) 'Monte Carlo tests with nuisance parameters: a general approach to finite sample inference and nonstandard asymptotics.' Mimeo, Department of Economic Science, University of Montreal.

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

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. (1986) Developments in the study of cointegrated economic variables, *Oxford Bulletin of Economics and Statistics* 48, 213-228.

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

Johansen, Søren (1988) ‘Statistical Analysis of Cointegration Vectors’, *Journal of Economic Dynamics and Control* 12, 231-54.

Johansen, Søren, (1991) ‘Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models’, *Econometrica* 59(6), 1551-80.

MacKinnon, J. G. (1991). ‘Critical values for cointegration tests,’ Ch. 13 of *Long-run Economic Relationships* (Engle and Granger, eds.) Oxford University Press 1991.

McLeod, A. I. and Li, W. K. (1983) ‘Diagnostic checking ARMA time series models using squared-residual autocorrelations’, *Journal of Time Series Analysis* 4, pp. 269-273.

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

Nordhaus, (1975) ‘Political Business Cycle’, *Review of Economic Studies* 42 (2),

Pissarides, C. (1980) ‘British Government Popularity and Economic Performance’, *Economic Journal* 90, 569-581

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

Schwarz, G. (1978) ‘Estimating the Dimension of a Model.’ *Annals of Statistics* 6, 461–4.