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

James Davidson*
Cardiff University

This version: March 2003

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 is 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. Bootstrap tests of the regular and generalized (non-)cointegration hypotheses are performed, as well as tests of the null hypothesis that cointegration of either type exists. Both the regular and double bootstrap statistics are calculated, the latter method providing a correction to the finite sample size distortion to the estimation of unknown parameters.

The tests reveal little or 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 case-specific Monte Carlo study, showing that the methods have ample power to reveal cointegrating relations, if they exist.

1 Introduction

A substantial literature has accumulated over recent decades, seeking theoretical and econometric links between economic conditions and the popularity of governments. Leading contributions include 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, 2002), Box-Steffensmeier and Smith (1996) and Dolado, Gonzalo and Mayoral (2001). Byers et. al. (1997), henceforth referred to as BDP, show that for the UK, 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 explain this finding by noting that opinion polls aggregate the heterogeneous opinions of voters who perceive economic circumstances differently, so that issues on whose significance voters are divided, even if important, may have little effect on support. Thus, borrowers and depositors take a different view of the interest and inflation rates. Likewise, the unemployment rate can mean different things to different people, witness the so-called 'North-South divide' and the contrasting fortunes of manufacturing and service industries, in the UK.

However, a formal test of the relationship remains wanting. Two statistical approaches to testing for time series relationships are in common use, the correlation approach and the cointegration approach. Neither of these is valid when the data in question are fractionally integrated. Since the party support series are 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 ($d > \frac{1}{2}$) this rate is n^{2d-1}, and the limit processes are not Brownian motion but fractional Brownian motion. The Brownian functionals that define the limit distributions depend on d, and the usual cointegration tests are inappropriate.

The present paper reports some tests of the cointegration hypothesis using the bootstrap, to overcome the problems with conventional tests. The theory of these tests is discussed at length in Davidson (2002a, 2002b). 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. Section 4 describes and reports bootstrap tests where the null hypothesis is noncointegration. Several variants of the null hypotheses are considered, including one in which the distribution of the bootstrap data under the null is based on the BDP model, allowing some data features to be captured that cannot be represented by a simple linear data generation process. On the other hand, Section 5 reports bootstrap tests where the null hypothesis is of cointegration. The consensus of the findings is that there is no discernible evidence against noncointegration, and only the most equivocal evidence in favour of cointegration. Section 6 gives the results of some Monte Carlo experiments designed to evaluate the power of these tests. Finally, Section 7 briefly reports a short-run correlation analysis, and Section 8 summarises the findings, and concludes.

---

[1]Since support measures are confined to the unit interval, the random walk is not, of course, a feasible model of the raw data. BDP model the series for $\log[\bar{X}_t/(1 - \bar{X}_t)]$ where $100\bar{X}_t$ is the sample average support. Since this process is defined on $(-\infty, +\infty)$ a random walk is a logically feasible representation. This would be manifested in the raw data by a tendency for support to cluster near either 100% or 0%, a phenomenon not commonly observed in democratic countries. In practice, note that 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.
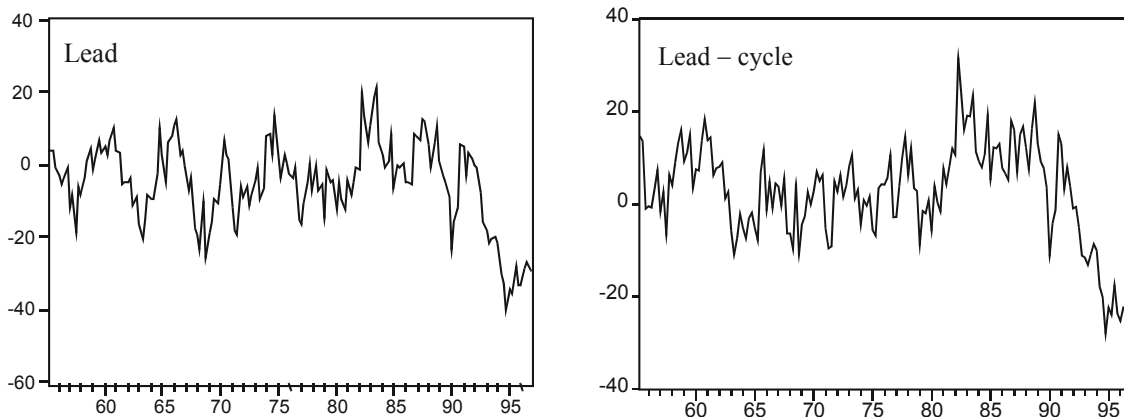
Figure 1: Lead, before and after removing the election cycle.

2   The data set

The data for the present study are quarterly observations for the period 1955:2 to 1996:4. The party support data are taken from the monthly Gallup poll series. The variable 'Lead' is measured as the end-of-quarter difference between Conservative and Labour percentage support in periods of Conservative government, and the difference between Labour and Conservative in periods of Labour government. This series is plotted in the first panel of Figure 1.

A set of dummy variables is used to represent the so-called 'election cycle' discussed in BDP. It has been observed that the popularity of governments, other things equal, depends on the proximity of the most recent and forthcoming elections, largely because of a tendency for voters to register a 'protest vote' in mid-term. In BDP, this effect is modelled as a quadratic function of the current government's elapsed term, and the effects were found to differ depending on whether Conservative or Labour is the party in power. To capture these effects, dummies are constructed as follows: (i) a zero-one 'Labour in power' dummy ('LabGovt'); (ii) the number of quarters elapsed since the last election ('Elapsed'); (iii) the square of (ii); (iv) the product of (i) and (ii); (v) the product of (i) and (iii). The second panel of Figure 1 shows the Lead series as residuals from the election cycle, fitted to the five dummies by least squares.

Six economic indicators, plotted in Figure 2, have been chosen as possible explanations of Lead. A valid test of the cointegration hypothesis requires the data be purged of deterministic trends, and a linear trend dummy is therefore included in the test equation. As an aid to intuition the 'detrended' variables (residuals from least squares regressions on constant and trend) are shown in Figure 3. These series all appear covariance nonstationary (see Table 2). After partialing out the various dummies, we posit the 'political business cycle' hypothesis as, in effect, the existence of a cointegrating relationship between the second series in Figure 1, and those in Figure 3. The null hypothesis of non-cointegration, by contrast, would imply either that some unmeasured 'non-economic' factors drive the variations in Lead, or (more plausibly) some variant of the BDP hypothesis.[2]

The results of running the full regression are shown in Table 1, and graphically in Figure 4.[3] The signs of several coefficients, such as unemployment and real earnings, are the opposite of what

---

[2] The possibility of a short term relationship in the differences also exists, and this is tested in Section 7.

[3] The usual standard errors and t statistics are reproduced for descriptive purposes, but of course their distribution is non-standard even asymptotically, so they cannot be used for purposes of inference.
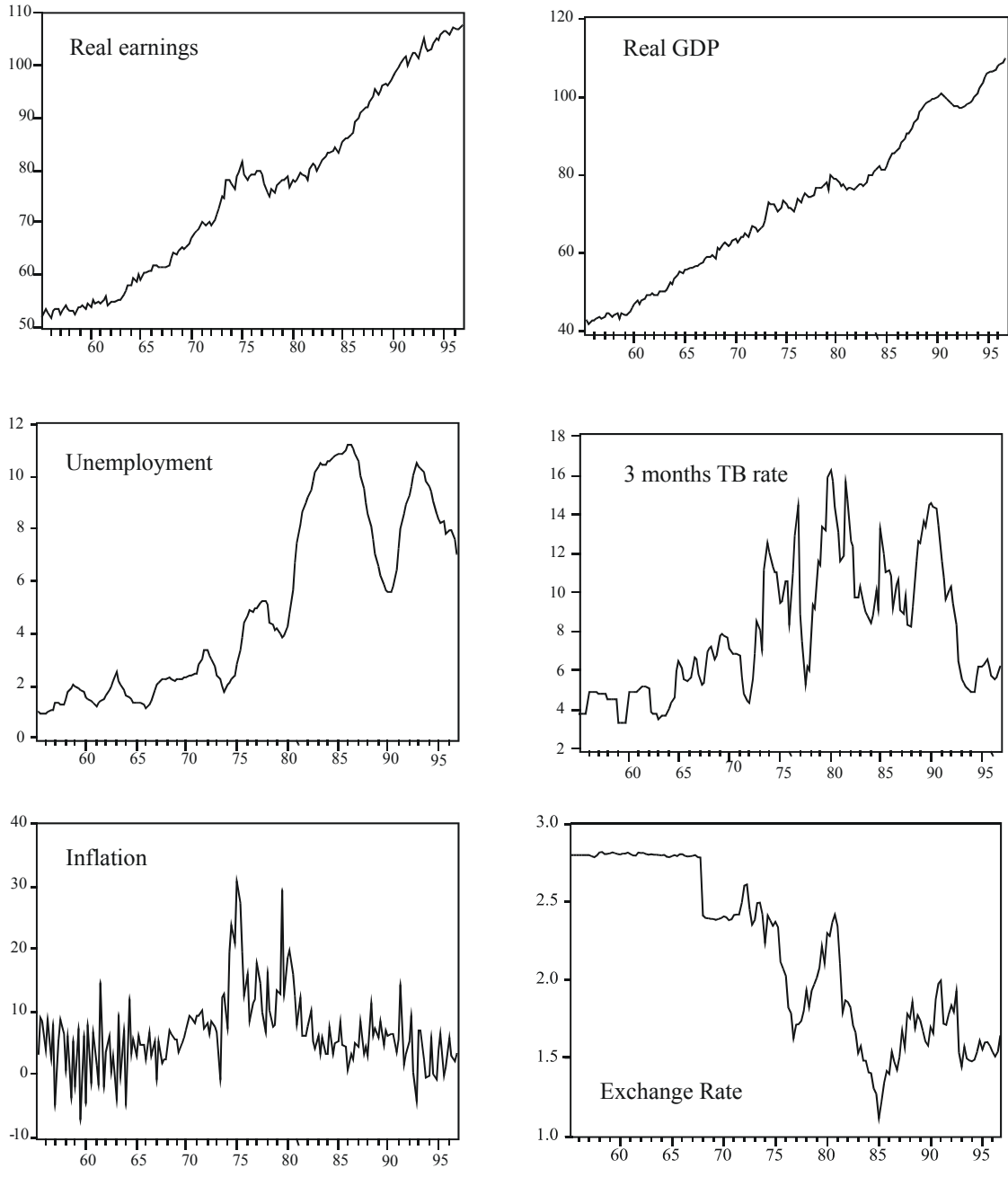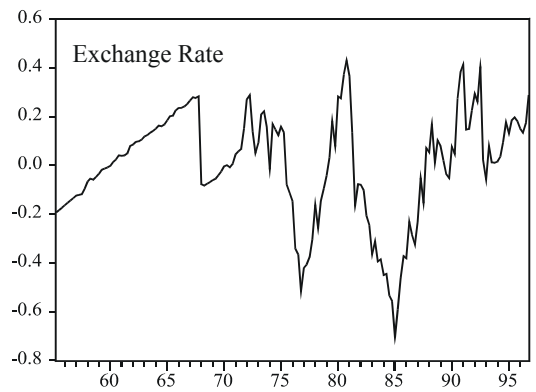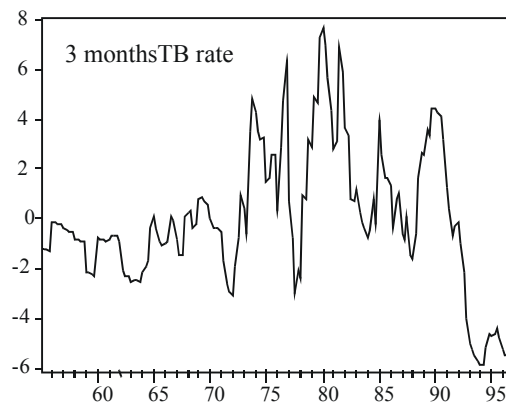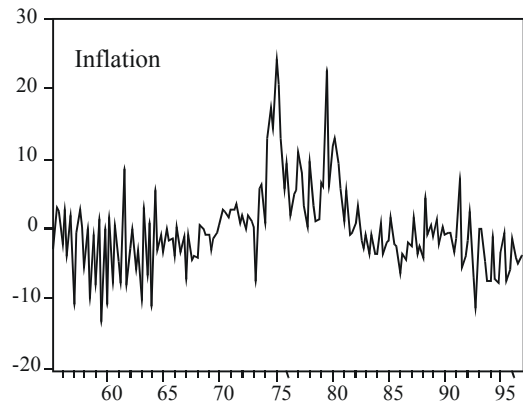
3

Figure 2: Economic Series
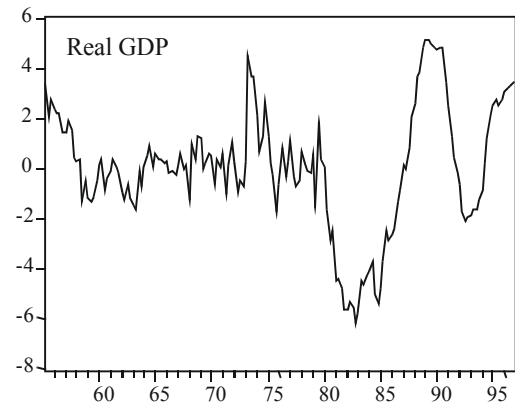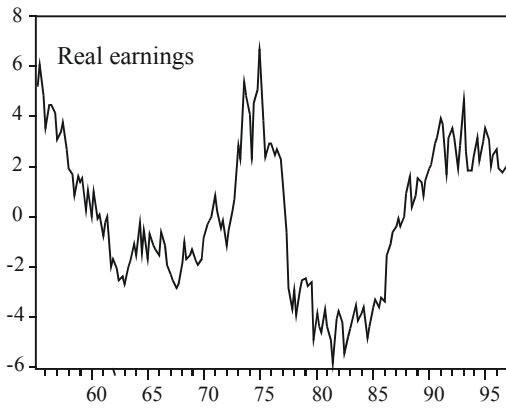
Figure 3: Series as deviations from trend

Dependent Variable: Lead   Sample: 1955:2 1996:4

	Coefficient	Std. Error	t-Statistic
Real Earnings	−0.509	0.335	−1.55
Real GDP	1.14	0.503	2.27
Unemployment	4.07	0.830	4.90
TB Rate	1.61	0.274	5.88
Inflation	−0.050	0.134	−0.375
Exch. Rate($/£)	1.53	4.537	0.337
LabGovt	12.31	3.53	3.48
Elapsed	−2.66	0.500	−5.32
Elapsed^2	0.141	0.027	5.06
LabGovt×Elapsed	−2.23	1.02	−2.17
LabGovt×Elapsed^2	0.086	0.060	1.41
Trend	−0.626	0.196	−3.19
Constant	−26.37	26.17	−1.00

R-squared	0.563	Adjusted R-squared	0.529
Durbin-Watson	0.789	F-statistic (all)	16.58
Residual PP	−6.381	F-statistic (econs)	19.12

Table 1: Regression of Lead on Economic Indicators

Figure 4: Regression of Lead on Cycle and Indicators

6

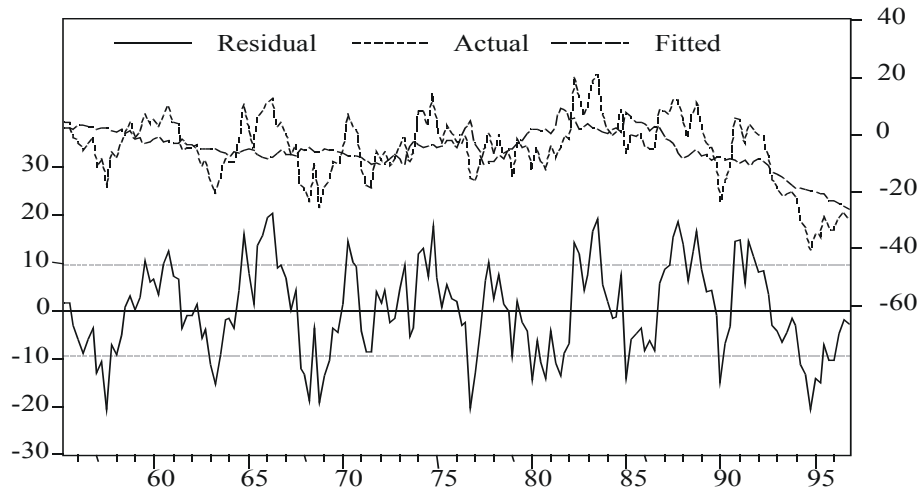|	Lead	RE	GDP	Unempl	TBR	Infl.	ExchR
d	0.765	0.920	0.978	1.169	0.626	0.664	0.991
	(0.066)	(0.079)	(0.060)	(0.150)	(0.107)	(0.092)	(0.108)
p	0	0	0	2	2	1	0
q	0	0	0	0	1	0	0
ARMA Coefficients:	-	-	-	0.518	0.397	0.473	-
				(0.16)	(0.105)	(0.135)	
	-	-	-	0.208	0.466	-	-
				(0.107)	(0.094)		
	-	-	-	-	0.814	-	-
					(0.038)		
Constant	0.084	46.59	38.98	-0.22	4.924	5.942	2.23
Trend	-0.061	0.353	0.401	0.062	0.037	0.007	-0.008
$Q(12)$ - levels	12.09	12.92	17.58	17.76	12.32	19.90	11.24
$Q(12)$ - squares	11.40	12.71	9.85	19.80	9.20	17.61	11.87

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

would be naively expected, although if the relationship turned out to be statistically significant these findings might need to be accounted for, rather than simply dismissed as spurious. Note that the residual Phillips-Perron statistic in Table 1, assuming six regressors and trend, 'rejects' nominally at the 5% level according to MacKinnon's (1991) tables. However the presence of the extra dummy variables are only the least of the complicating factors, in the correct interpretation this result.

Consider next the results of the univariate time series modelling exercises reported in Table 2. This table shows ARFIMA(p, d, q) models for each series in the data set,[4] chosen to optimise the Schwarz (1978) selection criterion, subject to the side condition that residual autocorrelation is insignificant by the Box-Pierce Q test for 12 lags. The second Box-Pierce statistic 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. This result may be compared with those obtained by BDP, who estimated the d coefficients for Conservative and Labour support separately, in monthly and quarterly data, and obtained values close to 0.7 in both cases.[5] Those authors also found that the estimated value was not sensitive to the removal of the election cycle. The difference of two I(d) processes is also I(d) in general, although we note that this series has the additional feature of being subject to occasional switches of sign. This occurs at election dates where government and opposition change places so that, in particular, Lead is negative at the relevant dates. Such switches occurred four times during the present sample period, in 1964.4, 1970.2, 1974.1 and 1979.2. Inspection of the chart in Figure 1 does not reveal very obvious jumps at these dates, but the fact that

---

[4]The ARFIMA estimates were computed by maximizing the Whittle likelihood, using the Ox 3.2 package available as Time Series Modelling 3.0; see Davidson (2003), Doornik (1999). The data are differenced to satisfy the stationarity/invertibility condition $|d| < 0.5$, and then 1 is added to the estimate of d so obtained. For the exchange rate model, the sample period excludes the fixed parity period 1955.2 to 1971.3.

[5]Fractional processes are asymptotically self-similar, so that the value of d does not depend on the frequency of the observations. The similar values obtained in monthly and quarterly support data accord with this interpretation.

the generation process has these nonlinear features should not be overlooked. We return to this question in Section 4.2.

3 Models of Fractional Cointegration

The findings in Table 2 pose some unexpected problems for the formulation and valid testing of the putative relationship. While unemployment, the interest and exchange rates and earnings all have estimated d insignificantly different from unity, this is not true of either Lead, or the interest and inflation rates. These variables 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 (2002b). Let this take the form

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

where

$$\boldsymbol{\Delta}(L) = \operatorname{diag}\{(1 - L)^{d_1}, \ldots, (1 - L)^{d_N}\} \tag{3.2}$$

$$\boldsymbol{K}(L) = \operatorname{diag}\{(1 - L)^{b_1}, \ldots, (1 - L)^{b_N}\} \tag{3.3}$$

where $d_1, \ldots, d_N$ are any nonnegative reals (assume $d_1 \geq \cdots \geq d_N$ without loss of generality), $0 \leq b_i \leq d_i$, $\boldsymbol{B}(L)$ is a $N \times N$ polynomial matrix 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. The model is modified here to include dummy variables $\boldsymbol{D}_t$ $(S \times 1)$, which in the present set-up include the time trend and the election cycle, with $\boldsymbol{\Phi}$ a constant $N \times S$ coefficient matrix. This system generates N series integrated to orders $d_1, \ldots, d_N$, such that

$$\boldsymbol{\Delta}(L)(\boldsymbol{x}_t + \boldsymbol{\Phi}\boldsymbol{D}_t) = \boldsymbol{w}_t \sim \mathrm{I}(0) \tag{3.4}$$

(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

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

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

$$\boldsymbol{\beta}'(\boldsymbol{x}_t + \boldsymbol{\Phi}\boldsymbol{D}_t) \sim \mathrm{I}(a). \tag{3.6}$$

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 (2002b), 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.

Dependent Variable: Lead	Sample: 1955:2 1996:4		
	Coefficient	Std. Error	t-Statistic
Real Earnings*	−.0821	0.505	−1.62
Real GDP*	0.788	0.836	0.942
Unemployment*	4.00	2.03	1.96
TB Rate	1.53	0.294	5.21
Inflation	−0.265	0.144	−1.83
Exch. Rate($/£)*	−11.4	5.28	−2.16
LabGovt	8.82	3.91	2.25
Elapsed	−2.85	0.562	−5.07
Elapsed^2	0.145	0.031	4.59
LabGovt×Elapsed	−2.57	1.17	−2.20
LabGovt×Elapsed^2	0.103	0.069	1.50
Trend	−0.156	0.054	−2.86
Constant	35.52	16.38	2.16

R-squared	0.420	Adjusted R-squared	0.375
Durbin-Watson	0.694	F-statistic (all)	9.31
Residual PP	−5.81	F-statistic (econs)	10.36

Table 3: The generalized cointegration model

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

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

This model has the peculiarity that the cointegrated series are not the elements of $\boldsymbol{x}_t$ themselves, but the fractional differences of orders $d_i - b$.[6] 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 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[7] to have a d of 0.765, based

---

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

[7]This means that the series have been transformed by the filter $(1 - L)^{d - 0.765}$ where d is the estimated fractional
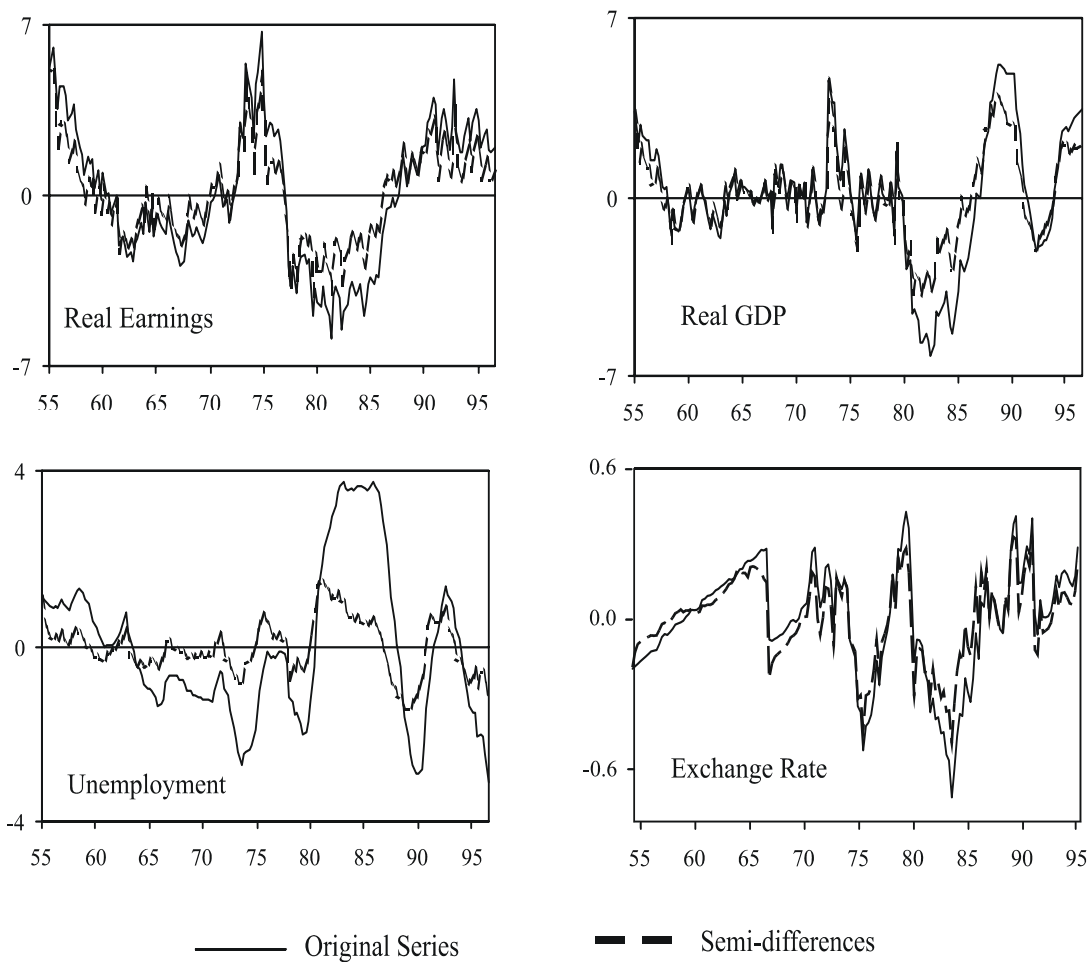
9

Figure 5: Semi-differenced series, $d = 0.765$

on the models in Table 2.[8] The filtered (and also detrended) series are shown, with the originals for comparison, in Figure 5. On the conventional criteria, this regression is somewhat inferior to the original in Table 1, and offers little support for the generalized cointegration approach in the present context.

___

integration parameter for the series in question. Note that the calculations are performed by truncating the expansions at lag t, for $t = 1, 2, \ldots$ See Davidson (2002b) for details.

[8]This model is actually a mixed case, in which the starred variables are assumed to make a generalized cointegrating relation with Lead for some $b \leq 0.765$, and then the unstarred variables further cointegrate with this set in the regular way to yield an I(a) residual, for $a < b$. With $a = 0.101$, this would be equivalent to the pure generalized cointegration model with $b = 0.664 = \min_i d_i$, but it also admits the stronger hypothesis in which $a = 0$. It's convenient for obvious reasons to keep Lead in its original form, and the bootstrap test has power to detect either case.

4 Tests of Noncointegration

4.1 The Basic Procedure

The bootstrap tests applied here are described in detail in Davidson (2002a, 2002b). The basic procedure is to draw bootstrap replications of the model in (3.1) under H_0, and so generate the null distributions of two regression-based test statistics. The actual statistics yielded by the regression in Table 1 are located in these empirical distributions to yield asymptotically valid p-values.

One of these statistics is the usual F statistic for the joint significance of the economic variables. This is not a conventional statistic to test for cointegration, but its consistency as a test for cointegration in the bootstrap context is proved in Davidson (2002b). Since it compares the sums of squared regression residuals with and without the test regressors included, which should have different orders of magnitude when the hypothesis of noncointegration is false, it should have some power to detect a relationship. The second statistic is the Durbin Watson statistic, which represents the more usual approach, of testing whether the regression residuals have significantly shorter memory than the variables constituting them. These therefore represent a two-pronged attack on the test problem, by looking at two different features of the fitted model. Being regression based, these tests may appear directly comparable with the Engle-Granger or Phillips-Perron residual-based tests, and this is true in the sense they can only test for the null of zero cointegrating rank.[9] However as we explain below, the tests may entail structural modelling of the short-run dynamics, and in this sense have much in common with system-based tests such as Johansen's eigenvalue tests.

The test statistics 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. Therefore, no conventional asymptotic tests can be based on them. However, this is also true of the more conventional tests. While there exist well known fixes to correct for such 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, so that a bootstrapping approach is unavoidable.[10]

The bootstrap draws are conditioned on the actual sample values of the regressors, which is computationally efficient and should also be less prone to specification error, because the generation processes of these variables do not need to be simulated[11]. This method also yields potentially more powerful tests than would bootstrapping the complete data set, noting that the conditional test distributions must have smaller dispersion than the unconditional ones.

Two versions of the null hypothesis can be distinguished, depending whether the matrix $\boldsymbol{B}(L)$ in (3.1) is assumed to be block-diagonal with respect to Lead (x_{1t}, say)[12] and the economic variables $\boldsymbol{x}_{2t}$, where $\boldsymbol{x}_t = (x_{1t}, \boldsymbol{x}_{2t}')'$. If it is, then the null under test is essentially that of independence

---

[9] Strictly speaking, the null hypothesis that is opposed to the regular cointegration alternative is that $\boldsymbol{\beta}$ contains no column in which Lead has a non-zero element. We have already noted that fractional cointegration amongst the I(1) regressors must be a feature of the alternative. Although the theory developed in Davidson (2002) deals with the simplest case of $\boldsymbol{\beta} = \boldsymbol{0}$, this generalization is perfectly valid provided the test regression is normalized on Lead.

[10] Breitung and Hassler (2002) show how to construct an asymptotically pivotal test of cointegrating rank by fractionally differencing the data, but their approach would be difficult to adapt to the present problem.

[11] ARFIMA models have been fitted to them in Table 2, but only for the purpose of modelling the short-run correlations of their increments with those of Lead, so that these can be incorporated into the conditional bootstrap model.

[12] Here we are ordering and partitioning the variables differently from equations (3.1), where the ordering was by size of d parameter.

between the series, which we call the strong null hypothesis. The bootstrap distribution of Lead would be simply obtained by using the univariate ARFIMA model from Table 2. In the case where this restriction is not imposed, which we call the weak null hypothesis, the short-run dynamics under H_0 have to be modelled to create the bootstrap distribution. Let $\boldsymbol{w}_t$ be the I(0) vector defined in (3.4). Because $\boldsymbol{x}_{2t}$ is 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 is to allow for the fact that w_{1t} could Granger-cause $\boldsymbol{w}_{2t}$, which is not ruled out, whether or not the regressors are weakly exogenous. With this structure, with leads/lags suitably chosen, the residuals from the regression should be asymptotically both serially uncorrelated, and orthogonal to the regressors at all orders. Resampling from the empirical distribution of these residuals, and then passing them back through the same filter in reverse, should accordingly yield a bootstrap sample having (asymptotically) the same correlation structure under H_0 as the original series. The resulting test distributions should therefore 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 will contaminate the short-run dynamics, and the best model must inevitably contain a large number of leads and lags. This problem is avoided only by choosing a deliberately parsimonious model, with short leads and lags, which should capture the weak dependence under H_0 but avoid contamination under the alternative. In practice, there is a trade-off of advantages between size and power. The simulations reported in Davidson (2002a) may throw a degree of light on the nature of this trade-off.

4.2   Modelling Lead

The simplest method of implementing the tests is to take the estimated value of d from the first column of Table 2, and use this to generate an I(d) series representing Lead, less the election cycle. The bootstrap series are non-cointegrated with the regressors by construction, but their increments reproduce the observed correlation structure with those of the regressors, under H_0, as described above. The use of this approach is supported in the present case by the fact that the Lead series is well described by a simple I(d) model, as is evident from Table 2.

However, this method has the drawback that it adopts at best a crude simulation of the process that is believed to generate the sample data under H_0. Note first that Lead must exhibit occasional jumps, corresponding to the sign change when governing and opposing parties change places. The marginal distribution of the innovations can be correctly simulated by bootstrap resampling on the actual model residuals, but this method cannot represent the state-dependence of the large deviations (the dates of elections are not randomly drawn) nor can it avoid the paradox of having 'Lead' remain negative both before and following an election. In other words, important nonlinear features of the data generation process have been lost. Moreover, by treating the election cycle as conditionally fixed, it ignores the fact that elections and 'Lead' are generated jointly. In the British political system, an election can be called by the governing party at any time up to the limit of the five-year term, and parliamentary terms of around four years are the norm, although they can be much shorter. The probability of an election being called ahead of time obviously depends on Lead itself, and is the less likely as Lead becomes negative.

An alternative approach is to construct the bootstrap distribution of Lead to embody these features. This has been done using the estimated BDP model as a basis, notwithstanding that this

is a monthly model fitted to a different, though largely overlapping, sample period. The procedure is as follows. First, monthly series for the log-odds of Conservative and Labour support, net of the election cycle, are generated by applying the appropriate fractional filter to the resampled BDP residuals. After adding the respective cycles from the BDP model to each series and converting to percentage form, the difference of the two series is constructed, signed according to the party in power, and finally, a quarterly series obtained by taking the figure for the last month in every quarter.

Within this setup, elections are modelled as random events whose probability depends on the number of unexpired periods of the current term, and also whether Lead is currently positive. This probability must in all events approach unity as the number of unexpired periods approaches zero. A simple scheme with the required properties is

$$P(\text{election in month } t) = \exp\{(\alpha - \beta \min(\text{Lead}, 0))(t - T)\}$$

where T denotes the last possible date of the current term. The parameters α and β are chosen by trial and error so that the number of elections, and changes of government, are typically close to those of the sample historical period. With $\alpha = 0.15$ and $\beta = 0.1$ the averages in 1000 replications were respectively 11.5 elections and 3.7 changes of government, which is close enough for the purposes of the exercise to the historical values of 11 and 4, respectively. Note that with this model, the dummy election cycle variables have to be resampled randomly, in each bootstrap replication.

This approach to doing the test has one drawback, that we cannot model the correlation between the increments of Lead and the economic indicators. In other words we are testing the 'strong' form of the noncointegration hypothesis, which is subject to the risk of spurious rejection, assuming we are only interested in the existence of a cointegrating relationship. There are three reasons why this is seen as an acceptable limitation; first, the strong hypothesis is in any case of independent interest, by throwing light on the short run; second, the evidence indicates that the amount of short-run correlation is small; and third, correcting the omission could not change the test outcome actually obtained.

4.3   Results

Table 4 shows the results of bootstrap tests for the four cases described above, in other words, the bootstrap models described respectively in Sections 4.1 and 4.2, applied to the regular cointegration model and the generalized cointegration model. The univariate ARFIMA models reported in Table 2 are used to provide estimates of the d parameters. To perform the generalized cointegration test the series for unemployment, real GDP and real earnings were semi-differenced as in Figure 5.

In the tests of the weak null hypothesis, based on the simple I(d) representation of Lead, its dth differences were modelled by regression on the lags of the regressors of lag orders -2 through $+2$, as well as 2 own-lags. This distribution was resampled, and passed back through the same filter in reverse. In the 'semi-weak null' the (not unreasonable) restriction that there is no feedback from Lead to the economic variables is imposed. In this case, future values of the differenced regressors are excluded from the dynamic model.

The bootstrap tests were performed with 1000 replications, although note that this number does not influence the precision of the estimated p-value. The tests are 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.

In the case of the weak and semi-weak null hypotheses, the reported p-values have been computed by the regular bootstrap and the double bootstrap. The double bootstrap has been

13

	F		DW	
Null Hypothesis	Regular	Double	Regular	Double
Regular (non-)cointegration:				
I(d), weak null	0.26	0.19	0.63	0.49
I(d), semi-weak null	0.31	0.27	0.59	0.42
BDP, strong null	0.12	-	0.69	-
Generalized (non-)cointegration:				
I(d), weak null	0.42	0.49	0.69	0.48
I(d), semi-weak null	0.46	0.38	0.65	0.50
BDP, strong null	0.26	-	0.76	-

Table 4: Noncointegration Tests: p-values

suggested by Beran (1988) as a method of minimizing size distortions. These are due to the fact that the bootstrap distribution differs from that of the sample data because it depends on estimated parameters. Such distortions are known to be worse when the statistics are not asymptotically pivotal, as in the present case. The method entails using the bootstrap p-value itself as the test statistic, since its asymptotic distribution under the null hypothesis is known (uniform on [0,1]). This is done by bootstrapping the bootstrap, a computationally intensive procedure that can, however, be made much more efficient by applying stopping rules, as proposed by Nankervis (2001).

The double bootstrap cannot compensate for errors of specification (as opposed to estimation) in the bootstrap model, which in this case means, pre-eminently, the incorrect choice of leads/lags in the weak-null specification. However, it has been noted that increasing the lag length must tend to reduce the probability of rejection in the event the null hypothesis is false. If the hypothesis is not rejected with the chosen model, as in the present case, increasing the lag length is not likely to change the decision. In the case of the BDP simulation the bootstrap parameters are extraneously estimated, so the double bootstrap is not available. However, note here that the parameters are computed from a much larger sample (monthly data) and moreover, the assumptions of the strong null are imposed so that no parameters estimated from the present sample are utilised. Since even this test fails to reject the null at the 5% level, it is difficult to see how changing the bootstrap specification could result in a reversal of the reported results.

In all these tests the p-values exceed 10%, so on this basis there is not even slender evidence of a cointegrating relationship. The set of economic indicators chosen may be incomplete, and for example the tax rate indicator used by Pissarides (1980) has 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.

5   Tests of Cointegration

Failure to reject the hypothesis of noncointegration at conventional significance levels may simply tell us something about the power of the tests. A natural next step is to interchange the null and alternative hypotheses, and see whether a rejection is obtained in this case. This can be done applying Shin's (1994) test for the null hypothesis of cointegration. Specifically, this is a test of the hypothesis that the residuals from the putative cointegrating regression are I(0), and while Shin derives it for the usual case of I(1) data, it can be adapted in the bootstrap context to the

	ECM ($\mu \leq -0.5$)		ECM ($\mu \leq -1$)		I(0) Regression	
Null Hypothesis	Regular	Double	Regular	Double	Regular	Double
Regular cointegration	0.018	0.076	0	0.005	0	0.001
Generalized cointegration	0.046	0.147	0	0	0	0.003

Table 5: Shin Test: p-values

case of I(d) data, as shown in Davidson (2002a).

The test is based on the KPSS statistic (Kwiatkowski et. al. 1992) from the cointegrating regression residuals that have been obtained using Saikkonen's (1990) efficient estimation procedure. The problem for adapting this test to the bootstrap context is to find a way of simulating the null hypothesis. In Davidson (2002a) this is done in two ways. The 'ECM' method is to fit an error correction model to the data. The bootstrap series for the dependent variable is then solved from this model with an inequality restriction on the error correction parameter. This restriction should be satisfied in the data if the null hypothesis is true, but in all cases will force the bootstrap data to be cointegrated, if the null is false. For the tests reported here, the upper bound on the ECM parameter has been set to the minimum of the estimated value and one of two bounding values, either -0.5, or -1. This choice represents a potential size-power trade-off, since making the bound smaller distorts the null distribution but must raise the probability of rejection when the null is false.

The second method of implementing the test is simply to regress the model residuals onto the fractionally differenced data, such that all the explanatory variables are I(0). Use of this model to construct the bootstrap series from resampled residuals again ensures that the null hypothesis holds for the series, although at the cost of some distortion of its distribution. For further details of these procedures, and formulae, see Davidson (2002a).

The results of these tests are shown in Table 5, where μ refers to the ECM parameter. The first columns show that the double bootstrap test with $\mu \leq -0.5$ cannot reject the null at the 5% level, although the other variants have no difficulty in doing so. However, the more relaxed bound must reduce the power of the test, in the interests of fixing the size more accurately. If the fitted value of μ in the sample data is much less that 0, this test cannot have much power, and this is a plausible occurrence even under the alternative, when the data are themselves fractionally integrated with $d < 1$, and the lag structure is potentially overfitted. In the present case the fitted value of μ is, respectively, -0.408 in the regular cointegration model, and -0.420 in the generalized cointegration model, whereas we know that these null hypotheses cannot both be true. The results of the next section also provide some grounds for placing most reliance on the other variants of the test.

6   Power Evaluation

6.1   Noncointegration Tests

One of the virtues of the bootstrap approach is that a power evaluation can be undertaken, relevant to the specific model under test. Consider first the tests of non-cointegration. Simulations have been used to estimate the power of the test of the weak and semi-weak nulls described in Section 4.1. Note that the "BDP strong null" cannot be evaluated in this way, simply because there is no way to construct a dummy alternative hypothesis.

The first part of the simulation procedure is to construct an artificial process representing a 'cointegrating residual', which is done using the fitted dynamic equation described in Section 4.1, so that the correlation structure of the differences in the observed data is preserved. Optionally,
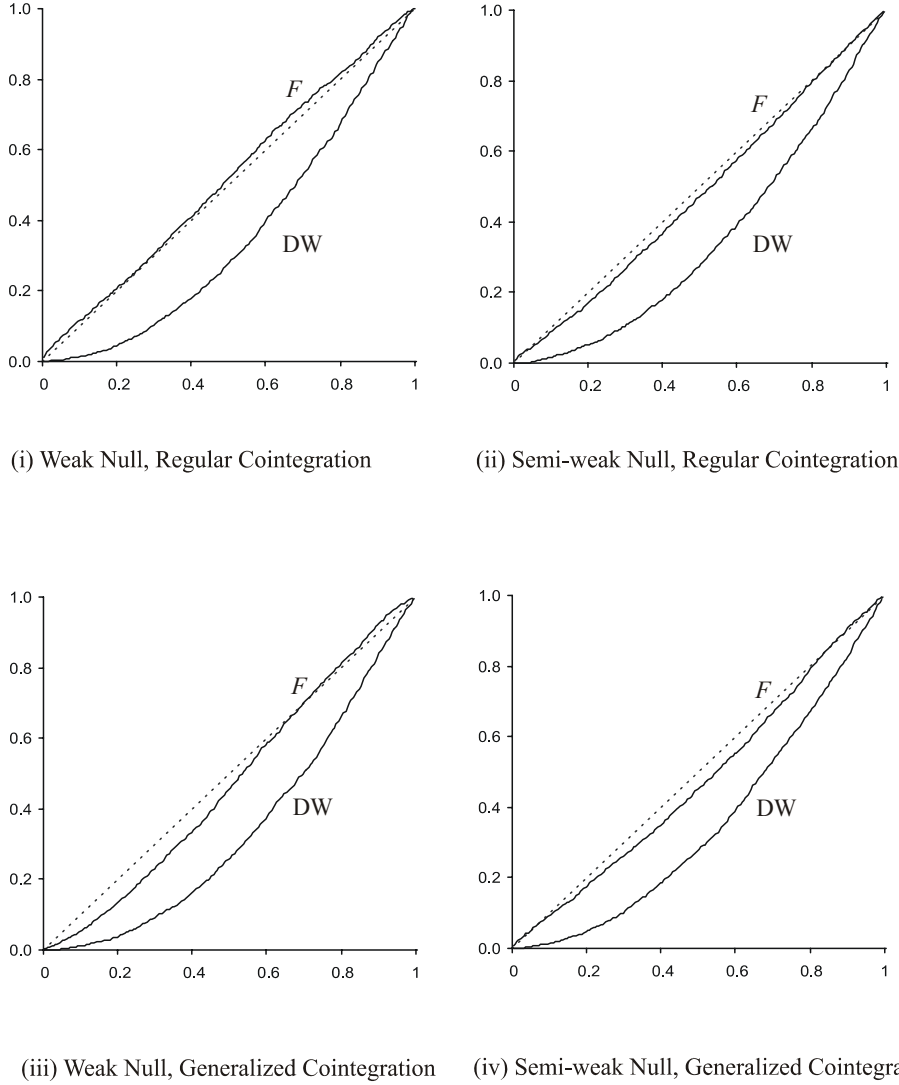
Figure 6: Empirical distribution functions of p-values for tests of non-cointegration.

this series is fractionally integrated, using an experimental d value. Finally, the fitted part of the regression in either Table 1 or Table 3 (to simulate regular or generalized cointegration, respectively) is added to the series, to produce an artificial regressand. When the experimental value of d is equal to 0, this method generates a fully cointegrated data set, while for cases $0 < d < 0.765$ the data may be called fractionally cointegrated.

An artificial 'null hypothesis' is created, in this framework, by integrating using the value $d = 0.765$ from Table 2, and then adding only the dummy components of the fitted model. This null model was simulated 3000 times to tabulate the true sizes of the tests. In each replication, the simulated test involves estimating the d value by maximum likelihood for the experimental dependent variable (the $\mathrm{ARFIMA}(0, d, 0)$ specification is used) as well as the parameters of the short-run dynamics. The empirical distribution functions of the p-values so obtained are shown in Figure 6. The EDFs of correctly sized tests, for which the p-values are $U(0, 1)$ by construction, should lie on the diagonals, shown by the dotted lines. These tabulations can be used to compute size-corrected p-values, a procedure that is effectively equivalent to the double bootstrap; see

16

		d	F		DW		R^2
			Nominal	Corrected	Nominal	Corrected	
		0.8	0.15	0.13	0.015	0.13	0.88
	Weak	0.6	0.36	0.33	0.043	0.25	0.93
	Null	0.4	0.65	0.63	0.18	0.54	0.95
		0.2	0.86	0.85	0.47	0.84	0.95
Regular		0	0.94	0.93	0.76	0.99	0.95
Cointegration		0.8	0.19	0.12	0.03	0.11	0.88
	Semi-	0.6	0.39	0.27	0.07	0.19	0.93
	Weak	0.4	0.64	0.52	0.19	0.45	0.95
	Null	0.2	0.85	0.77	0.50	0.79	0.95
		0	0.93	0.89	0.77	0.95	0.95
		0.8	0.08	0.15	0.007	0.12	0.80
	Weak	0.6	0.21	0.33	0.02	0.18	0.89
	Null	0.4	0.42	0.56	0.06	0.35	0.92
		0.2	0.69	0.80	0.19	0.68	0.92
Generalized		0	0.83	0.90	0.37	0.88	0.92
Cointegration		0.8	0.09	0.12	0.01	0.07	0.81
	Semi-	0.6	0.20	0.25	0.02	0.10	0.89
	Weak	0.4	0.44	0.49	0.07	0.21	0.92
	Null	0.2	0.71	0.74	0.20	0.44	0.92
		0	0.84	0.87	0.41	0.74	0.92

Table 6: Powers: Tests of noncointegration

Davidson (2002a) for details.

The test procedure was simulated 1000 times, for each of five values of the experimental d. The proportion of rejections at the 5% level are shown in Table 6. The two columns relating to each test show, respectively, the nominal powers, and the size-corrected powers, obtained by adjusting the p-values using the tabulations in Figure 6. The final column of the table shows the average of the 1000 R^2s obtained in the test regressions, to show how the rejection of non-cointegration relates to goodness of fit. Remember, in this connection, that the dummy variables for the election cycle account for a large proportion of the observed variation.

6.2 Shin Tests

For the power analysis of the Shin test, it is necessary to simulate the data under the null hypothesis of cointegration, to generate the EDFs for size correction. To do this the cointegrating residuals were generated using the data-fitted error correction model, modified by setting the ECM coefficient to the lesser of its sample value, and -1. The fitted part of the regressions in Tables 1 or 3 is then added to the generated residual to produce the artificial regressand. The resulting tabulations are shown in Figure 7.

There are two approaches to simulating an alternative hypothesis. The first is to generate the residuals as under the null, but then to fractionally integrate these using a spread of experimental d values (0.2, 0.4, 0.6, 0.8 and 1 are used) before adding the fitted components. The second, denoted UR in Table 7, is simply to replace the ECM coefficient by 0, so that the residual process contains a unit root. The former procedure is somewhat contrived, but it allows us to evaluate the power of the test of 'full' cointegration (with I(0) cointegrating residuals) against fractionally integrated alternatives.
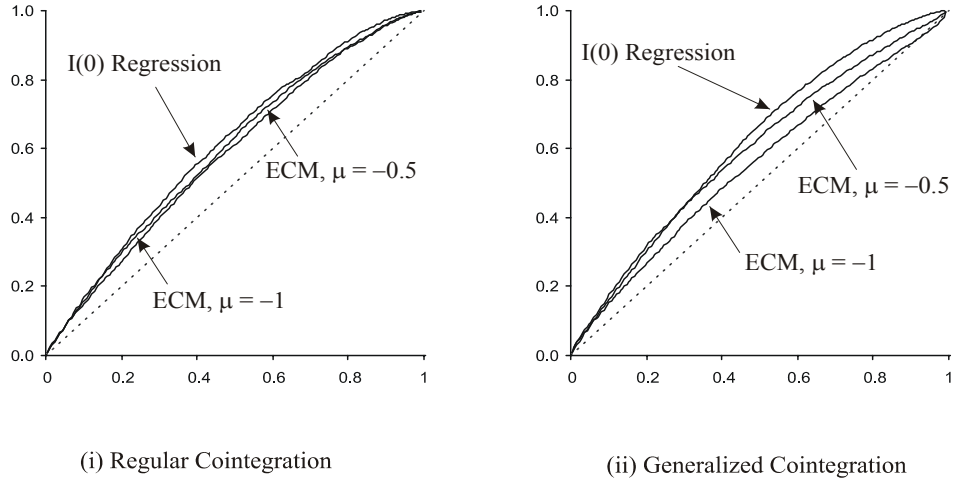
(i) Regular Cointegration

(ii) Generalized Cointegration

Figure 7: Empirical distribution functions of p-values for Shin tests of cointegration

	d	ECM, $\mu \leq -0.5$		ECM, $\mu \leq -1$		I(0) Regression		R^2
		Nominal	Corrected	Nominal	Corrected	Nominal	Corrected	
	0.2	0.19	0.14	0.20	0.14	0.35	0.24	0.94
Regular	0.4	0.33	0.24	0.45	0.37	0.72	0.61	0.94
Cointegration	0.6	0.56	0.47	0.76	0.70	0.93	0.88	0.91
	0.8	0.80	0.74	0.93	0.91	0.99	0.98	0.82
	1	0.93	0.91	0.99	0.98	1	0.99	0.76
	UR	0.78	0.73	0.97	0.96	0.98	0.97	0.82
	0.2	0.13	0.06	0.14	0.09	0.31	0.20	0.91
Generalized	0.4	0.29	0.17	0.38	0.30	0.69	0.55	0.89
Cointegration	0.6	0.52	0.38	0.66	0.59	0.93	0.87	0.82
	0.8	0.77	0.65	0.91	0.86	0.99	0.98	0.68
	1	0.90	0.82	0.99	0.98	1	0.99	0.72
	UR	0.78	0.66	0.97	0.94	0.99	0.98	0.68

Table 7: Powers: Shin Tests of Cointegration

6.3   Results

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. The actual data are used to represent the regressors in the simulations. While they should be generalized only with caution, the point is that, in principle, the same type of evaluation can be performed on a case-by-case basis. Second, note that the simulations of generalized cointegration do not tell us anything about the plausibility or otherwise of that particular model, but merely about the ability of the tests to detect (non)cointegration in fractionally integrated processes, where the common d is 0.765. In the Shin test, the cointegrated dependent variable under H_0 is I(1.169) (to match the largest regressor d) by construction. Hence, this evaluation is relevant to that state of the world in which the d for Lead has been mismeasured. The main contrast sought between these experiments is between the different orders of integration of the cointegrated variables.

Subject to these caveats, there is evidently reasonable power to detect both cointegration and non-cointegration, in this setting. The Durbin Watson-based test is shown in Figure 6 to under-reject quite seriously, but the F test is pretty well sized. The three variants of the Shin test all over-reject to a small degree, but the case with $\mu \leq 0.5$ appears not to be significantly better sized than the other two, even though there is a fairly substantial cost in terms of power. This finding allows us to be cautiously confident in relying on the clear rejections of the null, by the other two variants of the Shin test.

7   Tests for Short-run Correlation

There remains one further hypothesis that can be tested, that of a 'short term' relationship between the fractional differences. This is a possibility additional to, and distinct from, cointegration. BDP carry out a short-run analysis in monthly data by regressing the fractional differences of party support on innovations in a number of economic series, with generally negative results (see their Table 7). We have the opportunity here to carry out an analogue of the BDP tests in quarterly data, using the fractional differences of Lead as the dependent variable, instead of party support.

Two regressions were run. The first is on the I(0) series obtained by applying the operators $(1 - L)^d$ where d is the estimated differencing parameter from the relevant column of Table 2. In the second, the residuals from the ARFIMA(p, d, q) models were used, in those cases (Unemployment, TB Rate and Inflation) where there is a fitted ARMA component. In each case, the regressor set was chosen by optimising the Schwarz (1978) model selection criterion with respect to the number of included lags. In both regressions, zero lags (current values only) optimised the criterion. In the interests of space, just the F statistics for overall significance of the regressions are reported, together with the associated p-values; see Table 8, which also gives residual autocorrelation statistics.[13] The largest (absolutely) of the individual t values in these regressions is -2.19, on GDP, but note the perverse negative sign. These findings are quite closely comparable to the ones obtained in monthly data by BDP.

8   Conclusion

This paper has employed a variety of novel testing techniques to look for evidence of a connection between the popularity of UK governments and economic indicators over the business cycle, and

---

[13]The full regression results are available from the author on request.

	Significance of the Regression F(6,161) [Prob > F]	LM(4) Test for Autocorrelation F(4, 159) [Prob > F]
Fractional Differences	1.602 [0.149]	0.511 [0.727]
ARFIMA Residuals	1.552 [0.164]	0.552 [0.697]

Table 8: Tests for short-run correlation

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 spell out what conclusions can be drawn here.

First, in addition to the evidence on the 'short run' in Table 8, also note that the so-called 'strong' noncointegration hypothesis is effectively a test of statistical independence. Although these tests are not consistent against noncointegrating alternatives, the statistics will in general have their distributions shifted to the right in the presence of correlation between the process increments, even in the absence of cointegration.[14] Even in these cases, neither of the noncointegration tests reject, even at the 10% level, according to Table 4.

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 2000). Clearly, no such test can reject the non-cointegration hypothesis, and we can therefore treat these results as exact, such that the rejection probability under the null is known not to exceed the nominal level.

Third, the most hopeful result from the viewpoint of establishing the existence of a relationship is the nonrejection of generalized cointegration by the most conservative of the Shin test variants. However, this finding is at odds with the regressions reported in Tables 1 and 3, which reverse the ranking of the two models on the basis of goodness of fit. This anomaly points to the likelihood of a 'Type 2 error' in this case. The simulation evidence also gives us grounds to give credit to the more powerful test variants, which reject the null even after size-correction.

Fourth, while the set of economic series chosen for the test may omit some important ones, those included are undeniably important. The Monte Carlo evidence shows that the noncointegration tests have some power even against alternatives where the residuals are long-memory, and even nonstationary. The implication is that omission of important factors ought not, in general, to mask an existing relationship. If the important economic trend factors have been omitted, we are forced to the conclusion that these must be orthogonal to those included, and it is not at all obvious what these factors might be.

Fifth, in focusing on the formalities of cointegration testing we have not commented at length 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 Exchange Rate Mechanism. 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, this study can be claimed to provide, if anything, 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

---

[14]This property is demonstrated in Davidson (2002b).

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

Byers, D., J. Davidson, and D. A. Peel, (2000) 'The dynamics of aggregate political popularity: evidence from eight countries" *Electoral Studies* 19,1 49–62

Byers, D., J. Davidson, and D. A. Peel, (2002) 'Modelling political popularity: a correction' *Journal of the Royal Statistical Society* Series A 165, 187–189

Davidson, J. (2002a) 'Alternative bootstrap procedures for testing cointegration in fractionally integrated processes.' Working paper at http://www.cf.ac.uk/carbs/davidsonje

Davidson, J. (2002b) 'A model of fractional cointegration, and tests for cointegration using the bootstrap'. *Journal of Econometrics* 110, 187–212.

Davidson, J (2003) Time Series Modelling, Version 3.0, www.cf.ac.uk/carbs/econ/davidsonje/tsmod30.htm.

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.

Dolado, J., J. Gonzalo and L. Mayoral (2001) Long Range Dependence in Spanish Political Opinion Poll Series. *Journal of Applied Econometrics* (forthcoming).

Doornik, J.A. (1999), *Object-Oriented Matrix Programming Using Ox*, 3rd ed. London: Timberlake Consultants Press and Oxford: www.nuff.ox.ac.uk/Users/Doornik

Dufour, J.-M. (2000) '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.

Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y (1992) Testing the null hypothesis of stationarity against the alternative of a unit root. *Journal of Econometrics* 54, 159–78.

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

Nankervis, J. C. 2001, Stopping rules for double bootstrap tests. Working paper, University of Surrey.

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.

Saikkonen, P. 1991, Asymptotically efficient estimation of cointegration regressions, *Econometric Theory* 7, 1–21.

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

Shin, Y. 1994, A residual-based test of the null of cointegration against the alternative of no cointegration, *Econometric Theory* 10, 91-115.