Testing for Cointegrating Relationships with Near-Integrated Data

Suzanna De Boef
Pennsylvania State University and Harvard University

Jim Granato
Michigan State University

Testing theories about political change requires analysts to make assumptions about the memory of their time series. Applied analyses are often based on inferences that time series are integrated and cointegrated. Typically analyses rest on Dickey–Fuller pretests for unit roots and a test for cointegration based on the Engle–Granger two-step method. We argue that this approach is not a good one and use Monte Carlo analysis to show that these tests can lead analysts to conclude falsely that the data are cointegrated (or nearly cointegrated) when the data are near-integrated and not cointegrating. Further, analysts are likely to conclude falsely that the relationship is not cointegrated when it is. We show how inferences are highly sensitive to sample size and the signal-to-noise ratio in the data. We suggest three things. First, analysts should use the single equation error correction test for cointegrating relationships; second, caution is in order in all cases where near-integration is a reasonable alternative to unit roots; and third, analysts should drop the language of cointegration in many cases and adopt single-equation error correction models when the theory of error correction is relevant.

1 Introduction

Many of the questions political scientists seek to answer involve an understanding of the causes and consequences of political change. Testing theories about political dynamics requires analysts to make critical statistical assumptions about the memory of political time series. Intuitively, the memory of a time series refers to the rate at which the effects of shocks to a process—such as the effect of the massive changes in Congress in 1994 on the stream of policy outputs—dissipate. Memory may take many forms, but typically applied analysts consider only permanent memory (unit root processes), in which the stream of policy outputs is permanently altered, or short memory (stationary processes), in which the stream of policy outputs is briefly interrupted and quickly returns to preintervention levels. These assumptions about memory lie at the heart of our theories about political processes and they affect both the models we choose and the inferences we draw from them.

Authors' note: This research is based on work supported by the National Science Foundation under Grant 9753119. We thank Neil Ericsson for providing the experimental setups used by Kremers et al. (1992), which served as the basis for this analysis, and also for his continuing help and encouragement as we continue with this research project. Thanks also go to the CBRSS (Center for Basic Research in the Social Sciences) at Harvard University, Jack Medick for numerous helpful discussions, Mark Jones and Chris Achen for their comments, and Angel Toledo-Lopez for watching experimental results fly by on the computer screen.

Copyright 1999 by the Society for Political Methodology

In practice, political scientists often make two related claims about the memory of dynamic political processes. First, analysts claim that some time series data are integrated and specifically that they contain unit roots. Second, analysts often assert that two or more of these series trend together in the long run (i.e., that they are cointegrated). Durr (1993), for example, argues that domestic policy sentiment and economic fortunes are unit root processes which trend together because economic security is a necessary precursor to the acceptable implementation of expensive liberal policies. Conversely, downturns in the economy make expensive domestic policy less acceptable. Others assert that variables such as presidential approval and economic fortunes (Ostrom and Smith 1992; Clarke and Stewart 1996), levels of conflict and cooperation between countries (Rajmaira and Ward 1990), U.S. defense spending and public policy preferences (Wlezien 1996), and economic conditions and public support in parliamentary systems (Clarke and Stewart 1995; Clarke and Whitely 1997) are each unit root processes and that these pairs of variables trend together.

Taken together, these two claims—individual time series are integrated and jointly cointegrated—lead analysts to test hypotheses about political change using cointegration methodology. In general, political scientists use the Engle–Granger (1987) two-step methodology.[1]

The theoretical implications are direct. It is often argued that political time series are inconsistent with both the empirical and the theoretical properties of integrated time series. In this case, it makes little sense to theorize about integrated processes. Beck (1992) and Williams (1992), for example, point to the often restricted range of political time series, the absence of any kind of "growth" in political time series, and the tendency for mean reversion over long time periods as evidence that most political time series are not consistent with unit root processes.

In addition, analysts typically restrict decisions about memory to two choices: unit roots or stationary processes. De Boef and Granato (1997) and Box-Steffensmeier and Smith (1996) argue that restricting our understanding of memory to these two very different kinds of processes ignores long, but not infinite-memory, alternatives which may better describe our data. We know very little about the properties of estimators under these conditions.

On the other hand, this theoretical ambiguity is not easily resolved. While there are theoretical reasons to doubt that political data are integrated, statistical tests do not allow us to ascertain the memory properties of time series data with an acceptable level of certainty. Statistical tests for integration possess low power against local alternatives (Evans and Savin 1981, 1984; Phillips 1988; Blough 1992; De Boef and Granato 1997).

These arguments imply that we need to question the simple transference of cointegration methodology, particularly the Engle–Granger two-step methodology, to political science data and ask specifically what are the costs and benefits of applying cointegration methodology to data that are neither integrated nor stationary, but near-integrated.

We address some of these issues in this paper. Specifically, we assess the effects of nearly cointegrating relationships on the power of cointegration tests. Our assessment begins with a consideration of the power of pretests for unit roots, which are the basis for adopting cointegration methodology. The assessment continues with an evaluation of the power of tests for cointegration in the context of near-integrated processes which may or may not be cointegrating. Focusing on Dickey–Fuller (DF) tests from estimated cointegrating regressions and t tests from error correction models (ECM), we ask which test and which

[1] Alternative representations of cointegrating relationships are often used in economics. These include the Johansen vector error correction model (1988) and the Engle and Yoo three-step estimator (1991), but these techniques have not been widely used in political science. But see, for example, Granato and West (1994) or Clarke and Stewart (1994, 1995).

critical values analysts should use to distinguish better these alternative data generating processes (DGPs). We also examine the role of sample size and the risks associated with drawing inferences from each test under alternative conditions.

The paper proceeds as follows. In the next section, we discuss the concept of cointegration and its extension to near-integrated data. In Section 3, we present the DGPs for two simple and general processes. Next we review both the DF t tests for cointegration and the ECM representation of cointegrated variables with the attendant ECM t test for cointegration. In addition, we discuss the theoretical distributions of each test. In Section 5 we set up the Monte Carlo experiments and examine the results. Section 6 presents some examples of cases in which the two tests give different results. Finally, we discuss the practical implications of our findings for political scientists.

2 Cointegration and Near-Cointegration

Cointegration methodology is now commonly used in time series econometrics. Engle and Granger (1987) state that two or more series are cointegrated if each component series is integrated and some linear combination of these series is stationary.[2]

Formally, a process x_t has a unit root if in the autoregressive representation

$$x_t = \rho x_{t-1} + \mu_t \tag{1}$$

where $|\rho| = 1$.

For $|\rho| < 1$, the process is said to be stationary or I(0). The effect of shocks on a unit root process will accumulate into the future, while shocks to a stationary process will decay over time.[3]

Prior to Engle and Granger's (1987) contribution, the standard procedure for making integrated data stationary was differencing. But this type of transformation removes any long-run relationships in the data; differenced data reflect only short-run dynamics. If various data series share common trends or move together over time, these features are negated, and information is lost. This is not true when analysts use cointegration methods. Cointegration methods allow us to describe stationary equilibrium relationships between integrated series, preserving long-run information.

The popularity of cointegration methods, particularly error correction representations variously estimated, follows from other factors as well (Davidson et al. 1978; Banerjee et al. 1986; Hendry 1995). For one, cointegration methodology allows us to represent the data in a way that takes advantage of its theoretical properties. In particular, where it is argued (as above) that some variables "trend" together, cointegration methodology offers a close fit with theory. Further, Engle and Granger's two-step method matches intuition nicely with estimation and it is now widely used. The lack of parameter restrictions, the compatibility with theory, and the direct appeal of the Engle–Granger methodology make cointegration methodology an attractive alternative for political scientists analyzing the causes and consequences of political change.

Cointegration analyses begin with pretests for unit roots in the individual series of interest, typically using some form of the Dickey–Fuller (1979) test. Briefly, the series in question is first-differenced and regressed on its own lagged levels. If the coefficient on lagged

[2] This is a special case of cointegration and the one we consider.

[3] The mean and variance of a unit root process depend on time, so that the mean does not converge and the series' variance tends toward infinity. All references to integrated series in this paper refer more specifically to unit root processes.

levels is significantly different from zero, we reject the unit root null. If tests support the inference that the data are best characterized as I(1) processes, analysts typically adopt the Engle and Granger (1987) two-step methodology. In the first step, analysts estimate a cointegrating regression where levels of the dependent process are regressed on levels of the independent process(es).[4] The residuals from this regression are then also tested for the presence of a unit root using Dickey–Fuller tests.[5] If the null hypothesis that the series are not cointegrated (i.e., the residual series is nonstationary) can be rejected, then analysts proceed with step 2. Here the residuals from the cointegrating regression are entered into the second-stage error correction model (ECM) in which changes in the dependent process are regressed on changes in the independent process(es) and the previous period's equilibrium error (residuals from the cointegrating regression). This error correction model is then used to draw inferences about political change.

In spite of the attractiveness of cointegration methodology and the appeal of the two-step methodology, there are reasons to question the application of this methodology to the analysis of political change. There is some debate whether the properties of political time series are consistent with the theoretical and empirical properties of integrated data. In the limit, presidential approval, for example, cannot be integrated because its range is restricted to a unit interval and the series appears to cross some mean value with regularity. Similarly, the variance of the series is restricted and therefore is not strictly time dependent.[6] Alternatively, the properties of unit root processes may be mimicked by alternative long memory (Box-Steffensmeier and Smith 1996, 1998) or borderline processes (De Boef and Granato 1997) and these alternatives must be entertained whenever unit root processes are considered. The latter arguments are particularly relevant given the low power of unit root tests.

It has long been known that unit root tests have a low power against local alternatives (Evans and Savin 1984). Phillips (1987) developed local-to-unity asymptotics to formalize the logic in reference to local alternatives to unit roots. He called these local alternative processes near-integrated or near-I(1) processes. Phillips defined a near-integrated time series, $\{x_t\}$, as one for which the DGP has a root close to but not quite unity:

$$x_t = \rho x_{t-1} + \mu_t, \qquad \mu_t \sim I(0), \qquad |\rho| = 1 - c \tag{2}$$

where c is small. Near-integrated series are asymptotically stationary but behave as integrated series in finite samples (Banerjee et al. 1993; Phillips 1987). The range of c for which a series is near-integrated varies with the sample size. Shorter time series may be near-integrated for $c \leq 0.10$, while for longer sample periods, a time series with $c \leq 0.05$ or even 0.01 is required before the data mimic I(1) data and the sample series are near-integrated (see De Boef and Granato 1997).[7] While the effects of shocks on an integrated series never die out and the effects of shocks to a stationary series die out quickly, the effects of shocks to a near-integrated time series persist for some time. Further, in any given sample, distinguishing the extremely slow rates of decay of past shocks in a near-integrated

[4] The cointegrating regression may contain multiple independent variables as well as trends or events. In cases where the relationship is symmetric, the choice of which variable to treat as (in)dependent is irrelevant.

[5] The form of the test on the residuals from the cointegrating regression is the same as that used in the pretest phase; however, different critical values are used to reflect the fact that the residual series is estimated.

[6] For an exchange on this debate see Beck (1992), Williams (1992), and Smith (1992).

[7] Even longer time series may be needed to distinguish near-integrated prcesses from unit root processes when the series is estimated, as in the case of tests on the residuals from a cointegrating regression.

system from an absence of decay in an integrated system is impossible (Banerjee et al. 1993).

If pretests for unit roots have a low power against near-integrated alternatives, we must consider the possibility that tests for cointegration and estimated cointegrated models are based on near-integrated data. In these cases, the relationship may be near-cointegrating. Like series with explosive autoregressive roots, trends, or stochastic roots (Granger and Swanson 1996), cointegration may characterize relationships between near-integrated time series. We define these near-cointegrating relationships in a matter analogous to Engle and Granger (1987).

> Two or more series are near-cointegrated if each component series is near-I(1) and some linear combination of these series is I(0).

While asymptotically all linear combinations of near-integrated processes should be stationary, it is not clear that this will be true in finite samples. Our earlier work (De Boef and Granato 1997) demonstrates that many static regressions involving near-integrated data will find spurious relationships, similar to the unit root case. It is not clear, however, that cointegration methodology transfers in a similar fashion to near-integrated processes. There are several questions that need to be addressed before we can have confidence in inferences based on cointegration methodology as it has been applied in political science. Can tests distinguish cointegrating from near-cointegrating and noncointegrating relationships? Which tests and critical values perform well? Which have acceptable power and size? Ultimately, what are the properties of estimates of near-cointegrating relationships?

In this paper, we compare commonly used DF t tests based on cointegrating regressions with ECM t tests when data are near-integrated and near-cointegrated. We also consider the role of signal strength in drawing inferences. Signal–noise ratios have received little attention in political science, but their importance has been noted elsewhere (Kremers et al. 1992; Hansen 1995). Intuitively, the ratio represents the extent to which the error variances in the processes are similar. Larger ratios indicate that the error variance in the exogenous process is large relative to the endogenous process error.

3 Data Generation Processes

We consider a time series, $\{y_t, x_t\}$, which may or may not have a single cointegrating (or near-cointegrating) relationship. Specifically, we assume that the marginal process, x_t, is generated by an autoregressive process with a root (ρ) close to or equal to unity:

$$x_t = \rho x_{t-1} + \mu_t \tag{3}$$

with $\mu_t \sim \text{IN}(0, \sigma_\mu^2)$. Series such as presidential approval, macropartisanship, and economic expectations, for example, are often hypothesized to take this form.

The y_t DGP is a simple linear vector autoregression (VAR) with one-way Granger causality and weak exogeneity of x_t for y_t.[8] This parameterization is consistent with the models of partisanship proposed by Green et al. (1998) and of presidential approval proposed by

[8] The assumption of weak exogeneity is critical. It is also a reasonable assumption in many cases in which this type of DGP has been proposed in political science. Beck (1992) notes that in many cases the "symmetric treatment" of political time series does not make sense and that assuming weak exogeneity is reasonable. However, one should always test for weak exogeneity. Weak exogeneity cannot be tested directly, but tests for parameter constancy are the usual approach (Engle et al. 1983).

Ostrom and Smith (1992). It is also equivalent to the familiar autoregressive distributed lag model in the sense that the same relationship between the endogenous and exogenous variables is implied by each.[9] y_t is generated conditional on x_t and can be expressed as[10]

$$\Delta y_t = \alpha + \beta \Delta x_t + \lambda(y_{t-1} - x_{t-1}) + \varepsilon_t \tag{4}$$

Here ε_t is also IN(0, σ_ε^2) and independent of μ_t. Further, the ECM has a smaller error variance than the marginal process, $\sigma_\mu^2 > \sigma_\varepsilon^2$. The value of λ, the error correction rate, controls for the existence of a cointegrating relationship. If $\lambda = 0$, there is no cointegration and the error correction representation is invalid. If $\lambda < 0$, then the error correction representation is valid. Further, if the univariate series is near-integrated or integrated, the relationship is cointegrating (or near-cointegrating).[11]

4 Tests for Cointegration

Most tests for distinguishing a cointegrating relationship are based on the residuals from a cointegrating regression.[12] Political scientists typically estimate DF-type tests. As an alternative, Kremers et al. (1992) suggest a t test on λ in the ECM representation. We consider DF t tests of the null hypothesis of no cointegration and the ECM t test of the null of no cointegration.[13]

4.1 *Dickey–Fuller Tests*

When analysts follow the Engle and Granger two-step methodology, they typically conduct (augmented) Dickey–Fuller (1979) type tests at two stages. First, individual series are pretested to determine whether they are approximated by unit root processes. Second, they are used to test the null hypothesis of a unit root in the residuals of a cointegrating regression.[14]

[9] The ADL(1,1) model can be written as $y_t = \alpha + \phi y_{t-1} + \pi_0 x_t + \pi_1 x_{t-1} + \varepsilon_t$, where $\lambda = \phi - 1$, $\beta = \pi_0$, and $\pi_0 + \pi_1 + \phi - 1 = \alpha$ from the ECM representation. The DGP can thus be thought of as a process whose levels are caused both by its own past values and the current and past values of other variables. Alternatively, we can think of the DGP as a process whose changes are caused by past changes and the distance the series is from its long-run relationship (i.e., whether approval is too high or too low for current levels of economic conditions.).

[10] The true DGP imposes homogeneity, that is, $y_t = x_t$ in equilibrium. In practice, this assumption may be unreasonable. However, the results generalize in the absence of homogeneity. Note that the estimates of λ and β are unaffected whether we estimate $\Delta y_t = \alpha + \beta \Delta x_t + \lambda(y_{t-1} - x_{t-1}) + \gamma x_{t-1} + \varepsilon_t$ or $\Delta y_t = \alpha + \beta \Delta x_t + \lambda(y_{t-1} - \theta x_{t-1}) + \varepsilon_t$. DGPs without homogeneity can thus be similarly estimated by including an additional x_{t-1} term on the right (Banerjee et al. 1993).

[11] The ECM representation may be applicable when the data are stationary as well. Hendry (1995) elaborates on the versatility of ECM representations of stationary data.

[12] We might prefer to think of tests for cointegration in the context of near-integrated processes as tests to distinguish significant long-run relationships from spurious relationships rather than as tests for cointegration, but given our definition of near-cointegration, they are the same in finite samples.

[13] Other tests for cointegration include tests in the context of the Engle–Yoo three-step procedure (1991) and the Johansen (1988) procedure. If weak exogeneity appears to be an unreasonable assumption, then these tests take simultaneity into account. See Granato and West (1994) and Krause (1998) for applications of the Johansen procedure and Clarke and Stewart (1994, 1995) for use of the Engle–Yoo three-step procedure.

[14] It is important to note that the critical values for these test statistics vary given the presence of a constant or a trend in the cointegrating regression and the sample size. Further, these values are distinct from the critical values used when testing an individual series for a unit root. There is considerable controversy on unit root pretesting (Stock 1994; Elliott and Stock 1992). The critical values reported by MacKinnon (1990) are considered superior to those reported by Dickey and Fuller (1979) and Engle and Granger (1987).

Assume that univariate pretests lead us to conclude that the component series, y_t and x_t, are unit root processes.[15] We want to test whether the two series are cointegrated. Further, we have no expectations regarding the cointegrating coefficient, τ, relating levels of x and y in the cointegrating regression. In this case, we might proceed using ordinary least squares to estimate τ in a static regression:

$$y_t = \alpha + \tau x_t + \eta_t \tag{5}$$

Let $\hat{\eta}_t$ be the residuals from this regression. Now regressing $\hat{\eta}_t$ on its own lagged value, $\hat{\eta}_{t-1}$ without a constant yields an estimate of ϕ:

$$\widehat{\eta}_t = \phi \widehat{\eta}_{t-1} + \nu_t \tag{6}$$

The closer ϕ is to 1, the less likely the two series are cointegrated. The DF t test of the null hypothesis $\phi = 1$ is

$$t = \frac{(\hat{\phi}_t - 1)}{\hat{\sigma}_{\hat{\phi}_T}} \tag{7}$$

For large negative t, we reject the null hypothesis of no cointegration and $y_t - \alpha - \tau x_t = \eta_t \sim \mathrm{I}(0)$, with τ the cointegrating coefficient.

This is typically how the analyst proceeds. Alternatively, we may regress changes in $\hat{\eta}_t$ on lagged $\hat{\eta}_{t-1}$ (without a constant):

$$\Delta \hat{\eta}_t = \pi \hat{\eta}_{t-1} + \varepsilon_t \tag{8}$$

The closer $\hat{\pi}$ is to 0, the more likely ϕ is 1 and the less likely the series are cointegrated. In this case, the DF t test is now conducted using the familiar t test of the null hypothesis $\pi = 0$:

$$t = \frac{\hat{\pi}_t}{\hat{\sigma}_{\hat{\pi}_T}} \tag{9}$$

This representation is particularly useful for comparative purposes below.

4.2 *ECM Tests for Cointegration or Error Correction*

Dynamic estimates of the cointegrating coefficient(s) and the tests for cointegration based on these dynamic models outperform the DF tests based on static cointegrating regressions for some DGPs (Banerjee et al. 1986; Kremers et al. 1992). We consider the ECM t test for cointegration in this light. From (4) the ECM t test for cointegration is based on the t test for $\lambda = 0$:

$$\Delta y_t = \alpha + \beta \Delta x_t + \lambda(y_{t-1} - x_{t-1}) + \varepsilon_t \tag{10}$$

If $\lambda = 0$, then the distance between y_{t-1} and x_{t-1} is irrelevant for determining the nature of change in y: there is no error correction and therefore no cointegration. In contrast, if

[15] The pretest analyses proceed in the same manner as follows below for η.

this error correction term is significant, then future changes in y_t depend on whether y_{t-1} and x_{t-1} are close together or far apart. Testing the null hypothesis of no cointegration is thus the same as testing the null hypothesis that $\lambda = 0$. Kremers et al. (1992) recommend using MacKinnon critical values for the error correction model t test when the component series are individually I(1) processes.

4.3 *Comparing the Two Tests*

Kremers et al. delineate an important aspect of the differences between the two tests for cointegration. Let $w_t = (y_t - x_t)$.[16] Now subtract Δx_t from both sides of the ECM, rearrange, and substitute for $y_t - x_t$. We can now rewrite the ECM as

$$\Delta w_t = b w_{t-1} + e_t \tag{11}$$

where $\epsilon_t = (\beta - 1)\Delta x_t + \varepsilon_t$.

The t ratio on b for the ECM t test is exactly that on π from Eq. (8). Written this way, it is easy to see that the DF t test ignores the information in Δx_t or equivalently assumes $\beta = 1$. In other words, for the two tests to be equivalent, β must equal 1 or Δx_t must be 0. This will occur only if there are no short-term dynamics imposed by x_t. This "common factor" restriction in the DF t test is likely to be invalid and comes at some cost (Kremers et al. 1992).[17] The ECM t test makes no such restriction and thus uses more information than the DF t test.

The features of the distributions of the two t tests under the null hypothesis of no cointegration also suggest the dominance of the ECM t test.[18] The distribution of the DF t test is invariant to the dynamics omitted in the static regression: ($\epsilon_t = (\beta - 1)\Delta x_t + \varepsilon_t$). The distribution on the ECM t test depends on these dynamics. Specifically, as the variance of the error in the marginal process grows relative to the conditional process, the distribution approaches normal. Similarly, the more the short-run dynamics matter, the larger β, the smaller ϵ_t, and the more similar the distributions of the tests; the distribution approaches Dickey–Fuller. Taken together, these two parameters compose the signal-to-noise ratio,

$$q = -(\beta - 1)s \tag{12}$$

where $s = \sigma_\mu / \sigma_\varepsilon$.[19]

This dependence on the signal-to-noise ratio gives the ECM statistic a distinct advantage over the DF statistic when q is large and the component series are individually unit root processes. Features of the distributions of these tests will change under near-integration, but it is not clear how these changes affect results in finite samples. Monte Carlo experiments

[16] We assume that $\alpha = 0$ without loss of generality.

[17] To see the common factor restriction, express y_t in terms of w_t as $y_t = x_t + w_t$. Now given Eq. (11), rewrite w_t as $w_t = (1 + b)w_{t-1} + \psi$. Substituting into the equation for y_t, we have $y_t = x_t + (1 + b)w_{t-1} + \psi$. Finally, substitute for w_{t-1} lagged one period and collect terms: $y_t[1 - (1 + b)L] = [1 - (1 + b)L]x_t + \psi$, where L is a lag operator. Both y_t and x_t share a common factor in this representation, where the common factor is given by $[1 - (1 + b)L]$ (Hendry and Mizon 1978). This restriction is not made in the ECM t test. In other words, the ECM t test does not impose a common factor restriction.

[18] Under the alternative hypothesis of cointegration, or near-cointegration, w_t is a stationary process so standard asymptotic results apply. The distribution of both tests under the null hypothesis have been derived elsewhere and relevant results are presented in the Appendix.

[19] The derivation of s is given by De Boef and Granato (1999, Appendix, Part C).

provide insight into the size and power of these t tests when the component series are near-integrated.

5 Finite-Sample Evidence: Monte Carlo Experiments

We conduct two sets of Monte Carlo experiments. The first set considers the power of DF pretests for unit roots as the basis for performing cointegration tests. The second set of experiments examines the size and power of the DF and ECM t tests for cointegration when the series have unit roots or near-unit roots for both near-cointegrating and non-near-cointegrating relationships.

5.1 *Pretesting*

Conclusions from the ECM and cointegrating regressions are predicated on inferences drawn from unit root pretests. We wish then to consider values of ρ which are likely to pass the pretest phase. In prior work we found that in samples of 60 and less, values of $\rho \geq 0.9$ behaved as if they were integrated and, in particular, were prone to the spurious regression problem (De Boef and Granato 1997). We did not examine the performance of specific tests for unit roots.

We present estimated rejection frequencies of the null hypothesis that the component series are unit root processes for $\rho \leq 0.90$ and $p = 0.05$ using the DF t test in conjunction with critical values from MacKinnon (1990). These pretest results demonstrate the low power of the DF unit root tests as noted by Evans and Savin (1984) and Phillips (1988). The DGP for these simulations is given above in the marginal process (3) with $\mu_t \sim \text{IN}(0,1)$.[20] The form of the DF t test is given as above.

For each case, we generated 10,000 samples and estimated the DF test statistic. We recorded the rejection frequency for the null hypothesis that the process contains a unit root. The results indicate that in samples of size 40 for values of $\rho = 0.90$, we would reject the false null hypothesis only about 29% of the time and, for $\rho = 0.95$, in about 15% of the cases. For values of ρ closer to 1, the rejection rate falls to about 7%. These numbers are a very long way from the benchmark 95% rejection frequency. As the sample size increases to 60, the rejection frequencies are still quite low, but longer samples do increase the power of the test. The likelihood that an analyst would falsely conclude that near-integrated data are unit root processes is quite high and is never less than half, even when the form of the test statistic matches the DGP (see Table 1).[21] When a constant is incorrectly included in the test, the odds of false inference increase further.[22]

5.2 *Cointegration Tests*

We examine the size and power of the DF and ECM t tests for cointegration when the series have unit roots or near-unit roots for both near-cointegrating and non-near-cointegrating relationships. The DGP is as defined above in Eqs. (3) and (4). To allow for near-cointegration, we consider values of ρ in (3) ranging from 0.90 to 1.0, consistent with the above results.

We use 10,000 replications on the parameter space $q \times T \times \rho \times \lambda$, where $q = -(\beta - 1)(\sigma_\mu/\sigma_\varepsilon) = (0, 3, 8)$, $T = (40, 60)$, $\rho = (0.90, 0.91, \ldots, 1.0)$, $\lambda = (0, -0.05)$. This

[20] Rejection frequencies are unaffected by the variance of the error term.

[21] The results below should apply equally for larger samples and values of ρ closer to 1. For $\rho = 0.95$, we would falsely accept the null 50% of the time in samples as large as 130.

[22] We use PC-NAIVE (Hendry et al. 1990) to conduct the Monte Carlo simulations.

Table 1 Rejection frequencies: Null hypothesis, $\rho = 1$[a]

T	*0.90*	*0.95*	*0.96*	*0.97*	*0.98*	*0.99*	*1.00*
20	15.67	10.35	9.34	8.51	7.56	6.94	7.10
40	29.40	15.10	12.69	10.46	8.60	7.16	6.70
60	46.61	20.57	16.73	13.15	10.01	7.64	6.65
80	64.38	27.65	21.54	16.08	11.88	8.34	6.62
100	79.85	36.11	27.99	19.93	13.49	8.92	6.44
150	97.54	59.93	44.82	30.79	18.83	10.36	5.94
200	99.78	79.00	63.28	43.86	26.08	12.93	6.32

[a] Cell entries represent the rejection frequency of the null hypothesis that the series contains a unit root using the Dickey–Fuller test and MacKinnon critical values. The DGP is a simple autoregressive process with a root that ranges from 0.90 to 1.0 and no constant.

gives rise to 132 experiments for each test.[23] These values of q imply a wide range of signal-to-noise ratios.[24] We ought to see the largest differences in inference from the two tests when the common factor restriction imposed by the Dickey–Fuller test is most strongly violated. This corresponds to an experimental value of q equal to 8. In this case, the variance of the marginal process is quite large relative to that of the y_t DGP. We record the rejection frequencies for the null hypothesis $\lambda = 0$ as well as the DF rejection frequencies. We consider the MacKinnon and t critical values.

5.2.1 Results Under the Null Hypothesis

Under the null hypothesis that the series are not cointegrating, $w_t = y_t - x_t$ is a nonstationary process. It may be a unit root or a near-unit root process. In all cases in which the data do not contain cointegrating relationships, the ECM t test performs well, rejecting the null hypothesis with a great deal of power for all values of ρ and q (see columns 3–6 in Table 2). In particular, the MacKinnon critical values reject at, and in most cases well under, the 5% nominal rate for both sample sizes. The use of t critical values increases the odds of a type I error on average by a factor of under 2.

The rejection frequencies drop as the signal-to-noise ratio grows; the larger the error variance of x_t relative to y_t, the better the test is able to distinguish the null hypothesis. But differences are small across the experiments. In sum, the values of q and ρ have a negligible effect on inferences from the ECM t test when the null hypothesis is true: we are unlikely to reject the true null hypothesis that the data are not cointegrated or near-cointegrated.

The behavior of the DF t test is quite different, as the analytical results predict (see columns 7–10 in Table 2). Use of the DF t test produces results that vary greatly with T, ρ, q, and the chosen critical value.[25] First, consider the case in which $q = 0$. Kremers

[23] We fix the variance of σ_ε to 1 and vary σ_μ to simulate different values of q.

[24] For q to equal 0, β must equal 1. (In all other cases β is set to 0.5.) In this case, we chose unit variances so that the variance ratio, s, equals 1. Analytical results indicate that the statistics are invariant to s in this case. See the Appendix.

[25] Adding additional lagged values of changes in the residuals to pick up any autocorrelation in the DF t test, using the augmented Dickey–Fuller test (ADF), does not provide any significant improvement on the inferences drawn from the DF t test. This follows from the fact that the omitted dynamics in the DF t test are from x_t in the cointegrating regression itself, and not from lagged changes in the residuals. Thus the residuals from the cointegrating regression can be white noise even if the common factor restriction is not valid. As a result, adding

Table 2 Rejection frequencies for the true null hypothesis: No cointegration[a]

		ECM t test critical value				*Dickey–Fuller test critical value*			
		N = 60		*N = 40*		*N = 60*		*N = 40*	
q	ρ	*M*	*t*	*M*	*t*	*M*	*t*	*M*	*t*
8	0.90	2.75	5.71	3.09	5.55	39.20	57.22	25.16	39.74
8	0.95	3.28	5.63	3.30	5.57	18.20	31.98	13.93	22.49
8	0.98	3.30	5.32	3.32	5.66	9.72	17.15	8.03	13.69
8	0.99	3.33	5.16	3.37	5.48	7.19	12.64	6.46	11.25
8	1.00	3.24	5.27	3.35	5.46	5.08	9.09	4.89	8.88
3	0.90	3.54	6.39	3.43	6.10	23.78	36.81	17.33	27.26
3	0.95	3.97	6.78	3.78	6.46	14.83	24.80	11.98	19.26
3	0.98	4.00	6.59	3.89	6.43	9.72	17.15	8.03	13.69
3	0.99	3.75	6.47	3.70	6.32	7.08	12.62	6.46	10.95
3	1.00	3.75	6.43	3.74	6.20	5.34	9.24	5.30	8.93
0	0.90	4.81	8.80	4.87	8.41	4.94	9.06	4.95	8.60
0	0.95	4.83	8.86	4.98	8.51	4.94	9.06	4.95	8.60
0	0.98	4.99	8.89	4.99	8.57	4.94	9.06	4.95	8.60
0	0.99	4.91	8.90	5.01	8.62	4.94	9.06	4.95	8.60
0	1.00	4.91	9.02	5.03	8.70	4.94	9.06	4.95	8.60

[a] The MacKinnon (M) critical values are -1.9459 and -1.9493 for $N = 60$ and $N = 40$, respectively, for $p = 0.05$, one tail. The t critical values are -1.671 for $N = 60$ and -1.684 for $N = 40$. The signal-to-noise ratio, q, is given by $(\beta - 1)s$, where $s = \sigma_\mu/\sigma_\varepsilon$.

et al. (1992) note that since the distribution of the DF t test is invariant to the error variance when $q = 0$, the true value of $\beta = 1$. Thus when the experimental value of $\beta = 1$, the common factor restriction is valid and the two tests should coincide.

The evidence for $\rho = 1$ confirms this conclusion: the DF rejection frequencies for both sample sizes and all critical values are very similar to those obtained from ECM t tests. However, as ρ drops, the rejection frequencies using the DF t test remain constant while those using the ECM t test drop, even for $q = 0$. This implies that the actual distribution of the DF t test when the series are near-integrated is insensitive to ρ when $q = 0$. We cannot explain this somewhat surprising result. Further investigation is necessary.

The DF t test rejects the true null at higher rates than the ECM t test in virtually all cases as the signal-to-noise ratio grows. The differences between the tests grow as ρ falls. When ρ drops to 0.90, for example, even using the more conservative MacKinnon critical values, we overreject the null hypothesis at high rates for $T = 40$ and $T = 60$ (39.20 and 25.16%, respectively) when $q = 8$.

lagged changes in the residuals—using the augmented Dickey–Fuller test—will not help to solve the problem. We did repeat the experiments using the ADF t test to verify this claim. In none of the six sets of experiments does the ADF perform significantly better than the DF t test. At most, the true null hypothesis is rejected with a 2% lower frequency. This occurs with the smallest ρ and the larger sample—as intuition would suggest. When ρ is near 1, the rejection frequencies (using either set of critical values) differ by an average of less than 0.20. It is also the case that the test is more conservative whether the null hypothesis is true or false so that inferences will be neither clearly better nor worse using the ADF t test. Clearly, however, if the DGP is more general than that proposed here, the analyst should check the residuals from the DF t test to see if additional lags need to be included in the test.

The overrejection using DF t tests is not particularly surprising. Rejecting the null means rejecting a unit root in the residuals from the cointegrating regression. Even though pretests for unit roots in each component series indicate integration or near-integration (with these sample sizes and values of ρ), these series are not unit root processes. While pretests cannot distinguish I(1) from near-I(1) processes, it appears that the "posttest" on the linear combination more likely finds the residuals are stationary. This leads us to infer cointegration or near-cointegration when it does not exist, particularly as ρ drops, T increases, or q increases.

Using the Engle and Granger two-step method in conjunction with DF tests leads analysts to conclude falsely that the data are cointegrated, even with MacKinnon critical values. The problem is that the residuals from the cointegrating regression may be stationary even if the error correction representation is invalid, as long as the component series are not unit root processes. The inability of the DF t test to distinguish stationary and noncointegrating residuals from stationary and cointegrating residuals makes its use questionable.[26] The role of sample size is as expected. The more time points that are available for the test, the stronger evidence of decay indicating that the near-integrated component series are not unit root processes.

Importantly, if the signal-to-noise ratio is low, this is less true. For small q the residuals tend to look nonstationary; the true null hypothesis is not rejected even as $\rho \rightarrow 0.90$. In cases where the signal is noisy, but no cointegrating relationship exists, use of the Engle–Granger two-step methodology and the DF test suggests a false cointegrating relationship at unacceptably high rates. However, if the signal in the data is strong ($q = 8$), this danger is greatly diminished, particularly in small samples when posttests cannot distinguish near-integrated from unit root processes. Overall, the evidence supports the conclusion that the ECM t test, in conjunction with MacKinnon's critical values, performs better than the DF t test under the null hypothesis that the series are not cointegrated.

5.2.2 Results Under the Alternative Hypothesis

In tests where near-cointegrating and cointegrating relationships exist, the ECM t test should reject the null at high rates: 95% for $p = 0.05$. This is indeed true for high signal-to-noise ratios (see columns 3–6 in Table 3). Consider the results for $q = 8$. In this case, the power of the ECM t test is very high for all ρ and both sample sizes. For $T = 60$ the power is over 98% in all cases, regardless of whether one uses the MacKinnon or the t critical values (rejection frequencies using the t critical values are always slightly larger, but given that all are over 98%, the distinction is of little practical import). As the sample size drops, the test power remains high. The smaller sample size still allows for an efficient use of the information using the ECM t test. For $\rho = 1$, both critical values allow for rejection of the null of no cointegration at very high levels. These results are consistent with the fact that more observations provide more information with which to reject the false null hypothesis.[27]

The DF t test results have a much wider range of power, depending heavily on ρ and the critical values chosen (see columns 7–10 in Table 3). Consider the case where ρ is 1 and q is 8. DF tests relying on MacKinnon's critical values reject the false null hypothesis of no

[26] It may be that the absence of cointegration will be caught in the second step, but this is not known and we can provide no evidence in this regard.

[27] The results of Kremers et al. (1992), with $T = 20$, show a little less power—91.6 and 94.3% rejection frequencies for $p = 0.05$ for MacKinnon and t critical values, respectively.

Table 3 Rejection frequencies for the false null hypothesis: Cointegration[a]

		ECM t test critical value				*Dickey–Fuller test critical value*			
		$N = 60$		$N = 40$		$N = 60$		$N = 40$	
q	ρ	M	t	M	t	M	t	M	t
8	0.90	99.68	99.85	99.43	98.43	71.14	87.16	44.29	62.03
8	0.95	99.90	99.94	98.60	99.24	45.56	65.68	28.34	43.73
8	0.98	99.93	99.96	97.80	99.51	30.73	48.06	20.36	32.86
8	0.99	99.93	99.97	97.44	99.65	25.96	41.24	17.45	28.54
8	1.00	99.93	99.98	96.99	99.70	20.12	23.07	14.45	23.76
3	0.90	70.38	79.89	51.71	62.60	60.85	79.27	37.05	54.58
3	0.95	79.62	86.64	61.98	71.06	41.89	60.65	25.05	40.15
3	0.98	86.05	91.23	70.34	78.05	29.61	46.26	19.24	30.96
3	0.99	88.62	82.81	74.14	80.06	25.21	40.28	16.83	27.62
3	1.00	91.11	94.34	78.19	84.05	20.17	33.00	14.05	23.51
0	0.90	19.73	32.90	13.95	23.03	20.20	33.23	14.70	23.77
0	0.95	20.01	32.50	13.88	23.39	20.20	33.23	14.70	23.77
0	0.98	20.09	32.41	13.99	23.51	20.20	33.23	14.70	23.77
0	0.99	20.04	32.43	14.14	23.58	20.20	33.23	14.70	23.77
0	1.00	20.28	32.34	14.25	23.83	20.20	33.23	14.70	23.77

[a] The MacKinnon (M) critical values are -1.9459 and -1.9493 for $N = 60$ and $N = 40$, respectively, for $p = 0.05$, one tail. The t critical values are -1.671 for $N = 60$ and -1.684 for $N = 40$. The signal-to-noise ratio, q, is given by $(\beta - 1)s$, where $s = \sigma_\mu / \sigma_\varepsilon$.

cointegration just over 20% of the time when sample sizes are as large as 60, just over 14% when T is 40, and just 10% when $T = 20$ (Kremers et al. 1992).

Power is much improved using the larger t critical values. Notably, however, the test power is still quite low: often we will not reject the false null. As ρ drops to 0.90, the power improves. This is consistent with the increased likelihood that the residuals from the cointegrating regression are stationary: the raw series themselves will evidence faster rates of decay from shocks. It appears that even when the component series are near-integrated, we are more likely to conclude that the series are cointegrating than when the component series are integrated (using the DF t test).

However, as the signal strength drops (as q drops from 8 to 3), the rejection frequencies fall. Even for unit root series, the true alternative hypothesis is rejected in only about 80% of the cases for these mild signals in samples of 40 using the ECM t test. The situation worsens as ρ drops. In contrast, the DF t test results improve as ρ drops. This seemingly odd result is due to the fact that residuals from the static cointegrating regression are more likely to be stationary the smaller ρ, regardless of the existence of a long run, cointegrating relationship. Importantly, the DF t test still produces systematically smaller rejection rates than the ECM t test. As q drops further, neither test performs particularly well and use of either would result in similar incorrect inferences. Notably, the DF t tests are again invariant to ρ for $q = 0$.

By ignoring the dynamics, or imposing a common factor restriction, the DF t test performs poorly for high signal-to-noise ratios ($q = 8$), even for the larger sample size. This is particularly true for values of ρ near 1, but violating the common factor restriction also reduces the power and size of the test when ρ drops to 0.90. In contrast, the ECM t test performs as well as or better than the DF t test in all cases. For this reason, we recommend

that analysts use the ECM t test. We also suggest that analysts use the more conservative MacKinnon critical values.

6 Some Examples

The analytical results have distinct implications. When univariate tests indicate that time series are strongly autoregressive and near-integrated or that they are unit root processes, the DF and ECM t tests often lead to conflicting inferences. This is particularly true if the ratio of the square root of the error variances is large. We demonstrate this in the examples below. In these cases, our results indicate the superiority of the ECM t test.

We first consider monthly Republican macropartisanship and the Michigan Index of Consumer Sentiment (MICS) from 1981 through 1992, $T = 144$. We restrict the analysis to these 12 consecutive years of Republican presidents so that we may avoid the need for complicated transformations of the MICS necessitated by changes in the partisanship of the president. These series both wander within the sample period and cannot be said to show an affinity for a specific mean value. Theory suggests that both series are stationary. Shocks to these processes are believed to have short term effects on the level of the series, for example. Unit root tests on these univariate series show that each is strongly autoregressive and possibly I(1). Given these results, we might proceed to test for cointegration.

Using the DF t test based on the residuals of a static cointegrating regression of macropartisanship on MICS, we cannot reject the null hypothesis that the residuals contain a unit root ($t = -1.432$); the Engle–Granger two-step procedure indicates that cointegration is not present. While the pretests suggest that the series are integrated, no linear combination of the series is clearly stationary.

In contrast, the ECM t test is significant ($t = -2.259$), indicating the presence of an error correcting or cointegrating relationship. This result may be explained in part by the sizeable differences in the variation in the two series. Over this period, the standard error of macropartisanship is about 3 points, while that of MICS is over 11 points, suggesting a high signal-to-noise ratio. By omitting the dynamics, the DF t test ignores this information. The experimental results suggest that in this case the DF t test is likely to underreject the false null hypothesis and that inference should proceed based on the dynamic error correction model, assuming that all other assumptions of the model are satisfied. We would thus conclude that there is a long-run error correcting relationship between macropartisanship and consumer sentiment.[28]

There are also examples in which the ECM t test does not find a (near-) cointegrating relationship in the data, while the DF t test suggests that the relationship is cointegrating. Consider the relationship between liberal macroideology and inflation, quarterly from 1977 through 1994, $T = 72$.[29] Univariate tests indicate that the series are at least strongly autoregressive and may be unit root processes. Once again, the analyst might proceed to test for cointegration.

In this example, the DF t test from the cointegrating regression is significant ($t = -5.677$); we reject the null hypothesis that the residuals contain a unit root and infer that liberal macroideology and inflation are cointegrated. However, the ECM t test is not

[28] The ECM is estimated with an extra lagged independent variable to break homogeneity.

[29] Macroideology is measured as the percentage of respondents claiming to be liberal from the total number of respondents claiming to be either liberal or conservative. It may be thought of as the balance of liberal, relative to conservative, ideological self-placement.

significant ($t = -1.389$), leading us to infer that the data are not well represented by an error correction model and are not cointegrating.

Both macroideology and inflation exhibit much smaller variation than either macropartisanship or the MICS, with standard errors of 2.6 and 1.3 points, respectively. While the differences are much smaller, they suggest that the signal-to-noise ratio is still higher than 0. Appealing to the experimental results, we see that as the signal-to-noise ratio departs from 0, the size of the DF t test increases relative to that of the ECM t test. Once again, the experimental and analytical evidence suggests that inference should proceed based on the dynamic error correction model, assuming that all other assumptions of the model are satisfied. In this case, we conclude that there is no long-run cointegrating relationship between macroideology and the rate of inflation.

While each of these examples is based on underspecified models and we would not wish to draw any substantive inferences from them, it is clear that the use of DF and ECM t tests can lead to different inferences about the relationships between political time series. The analytical and experimental results point to the superiority of the ECM t test in these cases.

7 Conclusions

While De Boef and Granato (1997) conclude that near-integrated data are prone to the spurious regression problem, it is not clear that analysts can or should adopt cointegration methodology, at least not using the statistical tests traditionally employed to justify the analysis.

Near-cointegrating relationships are harder to identify than are cointegrating and non-cointegrating relationships. Specifically, DF t tests for cointegrating relationships used in conjunction with the cointegrating regression in the Engle and Granger two-step methodology are prone to find cointegration *when it does not exist* or to conclude that the relationships are not cointegrating when the true relationship *is* cointegrating. The tests have both a low power and a small size for cointegration. Further, when the power is low, the DF t tests have the wrong size.

Using the Engle and Granger two-step method in conjunction with DF t tests can lead analysts to conclude falsely that the data are cointegrated. The problem is that the residuals from the cointegrating regression may be stationary even if the error correction representation is invalid; the DF t test cannot make this distinction. It is possible that the absence of cointegration will be recognized in the second step—the second-stage error correction model will find no error correction—but this is not known and we can provide no evidence in this regard.

This issue as well as questions about the properties of the second-stage estimator are the subject of ongoing analysis (De Boef 1999). Banerjee et al. (1986) show that the estimates from the cointegrating regression are likely to be biased, particularly in small samples. It is not clear what "slippage" might result from the presence of near-unit root series, but it may be significant.

The role of sample size is as expected. The greater the number of time points available for the DF t test, the more likely the residuals from the cointegrating regression do not follow a unit root process. This is a double-edged sword. In cases where there is a cointegrating relationship we are more likely to find it with longer time series. However, when there is no cointegrating relationship, we are still more likely to find that the residuals from the cointegrating regression are stationary. *Thus larger sample sizes ensure better inferences under the alternative hypothesis but the opposite is true under the null hypothesis*.

The ECM t test performs reasonably well when the component series are near-integrated, particularly when there is no cointegration. In cases where the signal is strong, the ECM t test also performs well under the null. The same is not true, however, when the signal is relatively weak. In fact, when the signal is weak the ECM t test has no advantage over the DF t test.

While we have not compared alternative estimators of cointegrating and error correction relationships, this analysis informs us as to some of the merits of one step or single-equation error correction models. In applied analyses, ECMs of political change use the two-step estimator, but our results suggest that this may come at some cost.

Many of the relationships estimated using the Engle–Granger two-step methodology in applied research to date may be well represented by error correction models, but not necessarily be cointegrated or well estimated by two-stage estimators, or both. It may be true, for example, that domestic policy sentiment and economic fortunes move together because economic security is a necessary precursor to the acceptable implementation of expensive liberal policies, while downturns in the economy make expensive domestic policy less acceptable. It could also be the case that presidential approval and economic fortunes, the conflict and cooperation between countries, U.S. defense spending and public policy preferences, and economic conditions and public support for the British political parties and the prime minister, as well as Canadian party support, for example, are linked in the long run. These theoretical linkages seem reasonable.

The case where error correction itself is an invalid representation is much harder to find in applied analyses; most of our data are reasonably well represented by an ECM and may be estimated using a single-equation ECM. The single-equation ECM is known to have a solid empirical track record. This type of dynamic specification is a general reparameterization of an autoregressive distributed lag model in which levels of a process are a function of the series' own past and current values and past values of exogenous variables. In particular, the ECM imposes no (or less severe) restrictions on the DGP than do other transformations.[30] Indeed, the ECM can be estimated with stationary time series, and where the characterization of the univariate series is in doubt, the single-equation ECM has much to recommend it. This approach also has an additional implication: it is not necessary to use the language of cointegration to use the language of error correction.

Appendix: Test Distributions

The distributions of the test statistics are distinct under the null and the alternative hypotheses. Under the null hypothesis that the series are not cointegrating or near-cointegrating, $w_t = y_t - x_t$ is a nonstationary process. It may be a unit root or a near-unit root process. In either case, nonstandard asymptotics apply (see Hamilton 1994). Dickey and Fuller (1979) derived the distribution of the DF t test when the form of nonstationarity is a unit root process:[31]

$$t_{\mathrm{DF}} \xrightarrow{L} \frac{\int W_e dW_e}{\sqrt{\int W_e^2}} \tag{13}$$

[30] Assumptions about exogeneity become particularly important in the single-equation setup. For much of the literature in political science that relies on error correction representations, the necessary assumption of weak exogeneity is reasonable.

[31] All integrals range from 0 to 1.

W_e is the Wiener process for ϵ_t in Eq. (11). The distribution is skewed left, has a negative median, and is invariant to the error variance and β. The distribution resembles the standard t distribution as $w_t \rightarrow$ I(0). If the component series are near-I(1) [i.e., w_t is not I(1), but it is near-I(1)], the distribution is shifted by c and changes slightly in shape (Phillips 1987; Banerjee et al. 1993).

Kremers et al. (1992) define a normal approximation to the asymptotic distribution of the ECM t test for $\beta \neq 1$:[32]

$$t_{\text{ECM}} \xrightarrow{L} \frac{\int W_e dW_\varepsilon}{\sqrt{\int W_e^2}} \tag{14}$$

or

$$t_{\text{ECM}} \xrightarrow{L} \frac{(\beta-1)\int W_\mu dW_\varepsilon + s^{-1}\int W_\varepsilon dW_\varepsilon}{\sqrt{[(\beta-1)^2\int W_\mu^2 + 2(\beta-1)s^{-1}\int W_\mu W_\varepsilon + s^{-2}\int W_\varepsilon^2}} \tag{15}$$

where $s = \sigma_\mu/\sigma_\varepsilon$ is the ratio of the marginal process (3) to the conditional process (4).[33]

While the distribution of the DF t test is invariant to ϵ_t, the distribution of the ECM t test depends on ϵ_t. Recall that $\epsilon_t = (\beta-1)\Delta x_t + \varepsilon_t$. When $\beta = 1$, the distribution of the ECM t test is Dickey–Fuller because $\epsilon_t = \varepsilon_t$. For $\beta \neq 1$, we can write the distribution as

$$t_{\text{ECM}} \xrightarrow{L} \frac{\int W_\mu dW_\varepsilon - q^{-1}\int W_\varepsilon dW_\varepsilon}{\sqrt{\int W_\mu^2 + 2q^{-1}\int W_\mu W_\varepsilon + q^{-2}\int W_\varepsilon^2}} \tag{16}$$

where $q = -(\beta-1)s$. q gives the signal-to-noise ratio and depends on the extent to which the ratio of the square root of the error variances approaches 1 as well as on β. The distribution is sensitive to β and s only if they affect q.

For large q, the distribution of the ECM statistic is normal:

$$t_{\text{ECM}} \xrightarrow{L} N(0,1) + O_p(q^{-1}) \tag{17}$$

As q increases from 0, the distribution of the ECM statistic shifts from Dickey–Fuller to normal.[34] This gives the ECM statistic a distinct advantage over the DF statistic when q is large and the component series are individually unit root processes. As in the case of the DF t test, the distribution will be shifted and change shape as ρ drops.

Under the alternative hypothesis of cointegration or near-cointegration, w_t is a stationary process so standard asymptotic results apply for both tests.

[32] Banerjee et al. (1986) developed and Kremers et al. (1992) corrected the distributional results for the ECM t ratio.

[33] s may be thought of as a "nuisance parameter."

[34] O refers to the order of probability. A sequence of random variables is said to be $O_p(T^{-\frac{1}{2}})$ if for every $v > 0$ there exists an $M > 0$ such that $P\{|X_T| > (M/\sqrt{T})\} < v$ for all T (Hamilton 1994).

References

Banerjee, Anindya, Juan Dolado, David F. Hendry, and Gregor W. Smith. 1986. "Exploring Equilibrium Relationships in Econometrics Through Static Models: Some Monte Carlo Evidence." *Oxford Bulletin of Economics and Statistics* 48:253–277.

Banerjee, Anindya, Juan Dolado, John Galbraith, and David F. Hendry. 1993. *Cointegration, Error Correction, and the Econometric Analysis of Non-Stationary Data*. Oxford: Oxford University Press.

Beck, Nathaniel. 1992. "The Methodology of Cointegration." *Political Analysis* 4:237–248.

Blough, Stephen R. 1992. "The Relationship Between Power and Level for Generic Unit Root Tests in Finite Samples." *Journal of Applied Econometrics* 7:295–308.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90:567–580.

Box-Steffensmeier, Janet M., and Renee M. Smith. 1998. "Investigating Political Dynamics Using Fractional Integration Methods." *American Journal of Political Science* 42:661–689.

Clarke, Harold D., and Marianne C. Stewart. 1994. "Prospections, Retrospections, and Rationality: The 'Bankers' Model of Presidential Approval Reconsidered." *American Journal of Political Science* 38:1104–1123.

Clarke, Harold D., and Marianne C. Stewart. 1995. "Economic Evaluations, Prime Ministerial Approval and Governing Party Support: Rival Models Reconsidered." *British Journal of Political Science* 25:145–170.

Clarke, Harold D., and Marianne C. Stewart. 1996. "Economists and Electorates: The Subjective Economy of Governing Party Support in Canada." *European Journal of Political Research* 29:191–214.

Clarke, Harold D., Marianne C. Stewart, and Paul Whitely. 1997. "Tory Trends: Party Identification and the Dynamics of Conservative Support Since 1992." *British Journal of Political Science* 27:299–319.

Davidson, J. E. H., David F. Hendry, F. Srba, and S. Yeo. 1978. "Econometric Modeling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the United Kingdom." *Economic Journal* 88:661–692.

De Boef, Suzanna. 1999. "Modeling Near-Integrated Data," Working paper. Cambridge, MA: Harvard University.

De Boef, Suzanna, and Jim Granato. 1997. "Near-Integrated Data and the Analysis of Political Relationships." *American Journal of Political Science* 41:619–640.

De Boef, Suzanna, and Jim Granato. 1999. "Testing for Cointegrating Relationships when Data are Near-Integrated." Political Institutions and Public Choice (PIPC) Working Paper No. 98-08. East Lansing: Michigan State University.

Dickey, David, and Wayne Fuller. 1979. "Distribution of the Estimators for Autoregressive Time Series with a Unit Root." *Journal of the American Statistical Association* 84:427–431.

Durr, Robert H. 1993. "What Moves Policy Sentiment?" *American Political Science Review* 87:158–170.

Elliott, Graham, and James Stock. 1992. "Inference in Time Series Regression when the Order of Integration of a Regressor is Unknown." National Bureau of Economic Research Technical Working Paper No. 122.

Engle, Robert F., and C. W. J. Granger. 1987. "Cointegration and Error Correction: Representation, Estimation and Testing." *Econometrica* 55:251–276.

Engle, Robert F., and B. S. Yoo. 1991. *Cointegrated Economic Time Series: An Overview with New Results*. New York: Oxford University Press, pp. 237–266.

Engle, Robert F., David Hendry, and Jean-Francois Richard. 1983. "Exogeneity." *Econometrica* 51:277–304.

Evans, G. B. A., and N. E. Savin. 1981. "Testing for Unit Roots: 1." *Econometrica* 49(3):753–779.

Evans, G. B. A., and N. E. Savin. 1984. "Testing for Unit Roots: 2." *Econometrica* 52(5):1241–1269.

Granato, Jim, and William F. West. 1994. "Words and Deeds: Symbolic Politics and Decision Making at the Federal Reserve." *Economics and Politics* 6(3):231–253.

Granger, C. W. J., and Norman R. Swanson. 1996. "Further Developments in the Study of Cointegrated Variables." *Oxford Bulletin of Economics and Statistics* 58:537–553.

Green, Donald Philip, Bradley Palmquist, and Eric Schickler. 1998. "Macropartisanship: A Replication and Critique." *American Political Science Review* 92(4):883–899.

Hamilton, James D. 1994. *Time Series Analysis*. Princeton, NJ: Princeton University Press.

Hansen, Bruce. 1995. "Rethinking the Univariate Approach to Unit Root Testing." *Econometric Theory* 11:1148–1171.

Hendry, David. 1995. *Dynamic Econometrics*. New York: Oxford University Press.

Hendry, David, and Grayham E. Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance: A Comment on the Study of the Demand for Money by the Bank of England." *Economic Journal* 88:549–563.

Hendry, David, Adrian Neale, and Neil Ericsson. 1990. "PC-NAIVE: An Interactive Program for Monte Carlo Experimentation in Econometrics," Version 6.01.

Johansen, Soren. 1988. "Statistical Analysis of Cointegration Vectors." *Journal of Economic Dynamics and Control* 12:231–254.

Krause, George A. 1998. "Rivalry, Reciprocity, and the Dynamics of Presidential-Congressional Institution Building." Paper presented at the 1998 annual meeting of the American Political Science Association, Boston.

Kremers, Jeroen, Neil Ericsson, and Juan Dolado. 1992. "The Power of Cointegration Tests." *Oxford Bulletin of Economics and Statistics* 54:325–348.

MacKinnon, James. 1990. *Critical Values for Cointegration Tests*. New York: Oxford University Press.

Ostrom, Charles, and Renee Smith. 1992. "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." *Political Analysis* 4:127–183.

Phillips, P. C. B. 1987. "Towards a Unified Asymptotic Theory of Autoregression." *Biometrika* 74:535–548.

Phillips, P. C. B. 1988. "Regression Theory for Near-Integrated Time Series." *Econometrica* 6:1021–1043.

Rajmaira, Sheen, and Michael D. Ward. 1990. "Evolving Foreign Policy Norms: Reciprocity in the Superpower Triad." *International Studies Quarterly* 34:457–475.

Smith, Renee M. 1992. "Error Correction, Attractors, and Cointegration: Substantive and Methodological Issues." *Political Analysis* 4:249–254.

Stock, James. 1994. "Deciding Between I(1) and I(0)." *Journal of Econometrics* 63:105–131.

Williams, John T. 1992. "What Goes Around Comes Around: Unit Root Tests and Cointegration." *Political Analysis* 4:229–235.

Wlezien, Christopher. 1996. "Dynamics of Representation: The Case of U.S. Spending on Defense." *British Journal of Political Science* 26:81–103.