Long-Memoried Processes, Unit Roots, and Causal Inference in Political Science

John Freeman; Daniel Houser; Paul M. Kellstedt; John T. Williams

American Journal of Political Science, Vol. 42, No. 4 (Oct., 1998), 1289-1327.

Stable URL:
http://links.jstor.org/sici?sici=0092-5853%28199810%2942%3A4%3C1289%3ALPURAC%3E2.0.CO%3B2-%23

American Journal of Political Science is currently published by Midwest Political Science Association.

Your use of the JSTOR archive indicates your acceptance of JSTOR's Terms and Conditions of Use, available at http://www.jstor.org/about/terms.html. JSTOR's Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the JSTOR archive only for your personal, non-commercial use.

Please contact the publisher regarding any further use of this work. Publisher contact information may be obtained at http://www.jstor.org/journals/mpsa.html.

Each copy of any part of a JSTOR transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

JSTOR is an independent not-for-profit organization dedicated to creating and preserving a digital archive of scholarly journals. For more information regarding JSTOR, please contact support@jstor.org.

http://www.jstor.org/
Sun Sep 11 11:07:28 2005

Long-Memoried Processes, Unit Roots, and Causal Inference in Political Science

John Freeman, *University of Minnesota*
Daniel Houser, *University of Minnesota*
Paul M. Kellstedt, *Brown University*
John T. Williams, *Indiana University*

Theory: It has been argued that because researchers have not taken into account the long-memoried natures of certain political processes—especially the fact that some political time series appear to contain unit roots—some users of level Vector Autoregressions may have reached erroneous conclusions about the validity of important causal relationships and model specifications.
Hypothesis: For the first time, this argument is evaluated. The difficulties associated with modeling long-memoried political processes are reviewed. Then several approaches to dealing with them are discussed. One of the most promising approaches, Fully-Modified Vector Autoregression (FM-VAR) is studied in detail.
Method: The usefulness of FM-VAR is evaluated in a stylized Monte Carlo investigation and in reanalyses of major existing studies in political science—reanalyses that are representative of the ways in which level-VARs are employed in our discipline.
Results: Our experiments indicate that FM-VAR performs well (particularly in terms of size) in small and large samples, in fully and near-integrated systems, and in stationary systems. Most important, use of FM-VAR calls into question some of the major causal findings and specification test results in published studies. The implication, therefore, is that taking into account the trend properties of political processes is essential in theory building in political science.

The position that political theory often does not justify the strong restrictions which appear in our statistical models, but rather only weak restrictions, became well established in political science by the late 1980s (Freeman, Williams, and Lin 1989). In the years that followed, political scientists used this

This article has been through several revisions, largely due to comments we have received at professional conferences. Most recently, it was prepared for presentation at the Political Methodology Summer Meetings, Ann Arbor, July 18–21, 1996. An earlier version was presented at the annual meeting of the Midwest Political Science Association, Chicago, April 18–20, 1996. Freeman's work was supported by a grant from the Bank Austria Foundation. The authors thank Neal Beck, Suzanna DeBoef, Jim Granato, Mel Hinich, Yuichi Kitamura, George Krause, Mike MacKuen, Peter Pedroni, Janet Box-Steffensmeier, Renée Smith, and anonymous referees for their comments and suggestions. Neither these individuals nor the Bank Austria Foundation are responsible for the contents herein. The data and computer code necessary to replicate the findings herein can be obtained by contacting Paul Kellstedt at Paul_Kellstedt@brown.edu.

American Journal of Political Science, Vol. 42, No. 4, October 1998, Pp. 1289–1327 ©1998 by the Board of Regents of the University of Wisconsin System

methodology—VAR in levels, or level VAR—to evaluate competing theories in numerous fields (Goldstein and Freeman 1990, 1991; Freeman and Alt 1994; Williams 1990) and to assess the validity of the restrictions in their regression models (MacKuen, Erikson, and Stimson 1992). These evaluations were primarily based on Granger causality tests. The innovation accounting associated with VAR was used mostly for descriptive purposes.

At about this same time, econometricians identified and began to evaluate the problems that long-memoried processes like unit roots and cointegration produce in vector autoregression and related time series methods. These problems had to do with the validity of Granger causality tests which are at the heart of level VAR. This research was important because econometricians had discovered years before that many economic time series appear to be long-memoried (Nelson and Plosser 1982). In fact, economic theory predicts that certain time series will be random walks and that there will be no long-term relationships between them (see, for instance, Hall 1978; Juselius and Hargreaves 1992). Studying the trend properties of economic time series, therefore, became essential in time series econometrics.[1]

Taking note of these developments in economics, political scientists recently have argued that some of our time series processes are similarly long-memoried. Hence studies that employ level VARs could be in error. The first claim is based on both experimental and theoretical research—analyses of the trend properties of certain political time series, on the one hand, and derivations from rational-choice models of politics on the other (Ostrom and Smith 1993; Durr 1993a; Williams and McGinnis 1988; Chrystal and Peel 1986). The second claim is more conjectural. Scholars have speculated that by failing to account for the unit roots or, more generally, the trend properties of some political data in our level VARs, we may have drawn incorrect inferences about the validity of competing theoretical claims and misspecified our regression models (Granato and Smith 1994). Yet, to date, no political scientist has made any such demonstration, let alone explained what should be done to put our results on sounder footings if, in fact, some of our level VARs are faulty.[2]

[1]The main distinction in terms of trend properties is that between difference and trend stationarity. The former is associated with unit root processes while the latter with deterministic time trends. See, for instance, Hamilton 1994 (Chapter 15) and such applied works as Nelson and Kang 1981. Some branches of macroeconomics employ nonlinear conceptions of trends. Real business cycle theory is one such branch (see Freeman and Houser 1998). The focus in this paper is on causality tests. But as new work in econometrics (Phillips forthcoming) shows, unit roots also have important implications for the innovation accounting associated with VAR. See note 5.

[2]It is worth noting that, in some cases, theory implies political processes that are stationary. Hence, in those cases, the problem studied here may not apply. See, for example, the exchange between Smith (1993), Williams (1993), and Beck (1993). However, the experimental studies cited in the text indicate that, in many other cases, political time series are well approximated by random

This paper is the first demonstration of this kind applied to political science. In it, we explain the problems that unit roots and cointegration produce in level VARs—why it is so important to take into account the trend properties of one's data. We then review several approaches to solving these problems. One of these approaches, the Phillips (1995) Fully Modified Vector Autoregression (FM-VAR), is singled out for closer study. The nature of FM-VAR is explained, and some practical difficulties in implementing the associated estimation techniques and hypothesis tests are discussed. Finally, the usefulness of FM-VAR is explored in a stylized Monte Carlo analysis and in several analyses that parallel the main uses of level VARs in political science. The latter are a reanalysis of the Freeman (1983) study of arms races and a retest of the specifications of the MacKuen, Erikson, and Stimson (1992) model of approval.

The results are very enlightening. The Monte Carlo investigation shows the FM-VAR Wald test statistic performs well in terms of size and power in small and large samples in fully and near-integrated systems as well as in stationary systems. Using this statistic, we confirm Freeman's (1983) finding that Indian arms spending Granger causes Pakistani arms spending in the period 1948–1975; provision for theoretically expected unit roots in the respective series (Williams and McGinnis 1988) leaves Freeman's basic finding unchanged. A second reanalysis reveals potential problems with MacKuen, Erikson, and Stimson's (1992) study of the sources of presidential approval. Application of FM-VAR shows evidence of simultaneity between approval and business expectations. This calls into question MacKuen, Erikson, and Stimson's use of recursive models in modeling approval. To be more specific, the existence of unit roots and cointegration in some of those authors' time series casts doubt upon their specification test results (and hence model inferences as well). The implications of these findings for theory building in political science are discussed in the conclusion of the paper.

1. The Problem

The view that much political theory produces only weak restrictions on statistical models is akin to the idea that, at most, we can analyze the reduced forms of what are in reality unknown structural-equation models of politics. Much political theory is a loose collection of causal claims, claims that are best assessed through causality testing and innovation accounting in the context of a level VAR, a model that makes comparatively weaker assumptions

walks with infinite variance (just as in physics, investigations show that Brownian motion usefully approximates processes that do not have infinite energy; Melvin Hinich, personal communication). Similarly, the formal-modeling papers cited in the text show rational political agents of various types behave like random walks (just as rational-expectations theory does in economics).

than the conventional structural-equation models (see Freeman, Williams, and Lin 1989, especially 853–5; see also Hamilton 1994, 326–7).

When there are no stochastic trends in the data, the procedures for correctly estimating vector autoregressions are well known. That is, if the errors are serially uncorrelated, OLS yields consistent estimates of the coefficients, and familiar distribution theory is justified asymptotically. Also, certain methods are available to deal with such problems as time series outliers.[3]

The problem is this: If a level VAR contains a process that has a unit root (or a cointegrated relationship), it will not be a reduced form. And causality tests based on OLS estimation of such a level VAR can yield results which are second-order biased. In particular, the limit distribution of the statistics for tests of the joint statistical significance of combinations of coefficients—some of which are coefficients of the nonstationary variables—may have nonstandard shapes (that is, fatter tails than the standard distributions). If this fact is not taken into account and standard distributional assumptions are made, the respective causality tests may yield mistaken inferences.[4]

1.1 Formal Statement

The problem can be explained formally as follows.[5] Suppose one fits a first order n-vector autoregression of the form:

$$y_t = Ay_{t-1} + \varepsilon_t \tag{1}$$

where ε_t is iid$(0,\Sigma_{\varepsilon\varepsilon})$, $\Sigma_{\varepsilon\varepsilon}$ is p.d., and the initialization y_0 is any random n-vector. Suppose further that A in Equation 1 has the form

[3]On the estimation of single-equation autoregressions, see Hamilton (1995, Section 8.2, case 4). On the parallel multi-equation (VAR) estimations see Hamilton's Chapter 18. For solutions to the problem of time-series specification in the presence of outliers see Martin 1980, Tsay 1986, and Le, Raftery, and Martin 1996; see also Kitagawa 1987.

[4]First-order bias has to do with familiar conditions of consistency; second-order bias is defined in terms of the location and shape of the limiting distribution of estimators. Denote the variables by y_t, the errors by e_t, and let T be the sample size. Then it is true by construction that the expectation of e_t given y_{t-1} is zero. However, if there is a unit root or cointegration in the level VAR, plim $T^{-1} \Sigma y_{t-1}e_t$ is not zero.

[5]What follows is a condensation and slight elaboration of Phillips 1995, Section 2. See also Phillips 1992a,b. Following Phillips, we use "simultaneity bias" and "endogeneity." (See note 4.) Again we focus here on causality tests. But in a new paper, Phillips (forthcoming) shows that, for VARs with roots at or near unity and some cointegration, the associated innovation accounting is problematic. In particular, the long period ahead impulse responses are inconsistent; as the sample size grows, the estimated impulse response becomes a random variable rather than the true impulse response. In addition, forecase error variance decompositions for such VARs are inconsistent. Failure to account for (near) unit roots and cointegration can result in estimates of the forecast error variance at long horizons that are in error; usually one will underestimate this variance.

$$A = \begin{bmatrix} 0 & B \\ 0 & I_{n-r} \end{bmatrix} = (A_{ij})$$

where B is an r by (n-r) matrix and I_{n-r} is the corresponding (n-r) identity matrix. This suggests a partitioning of the original y_t vector, $(y'_{1t}, y'_{2t})'$ so that Equation 1 can be rewritten

$$y_{1t} = By_{2t-1} + \varepsilon_{1t} \tag{1a}$$

$$y_{2t} = y_{2t-1} + \varepsilon_{2t}. \tag{1b}$$

Thus y_{2t} is a full rank I(1) process, and y_{1t} is cointegrated with y_{2t}. In other words, the simple first-order level VAR in (1) is a system in which there are (n-r) unit roots and r cointegrating vectors. The latter have the form $\beta' = [I_r - B]$. Premultiplication of Equation 1 by β' gives the stationary relationship:

$$\beta' y_t = y_{1t} - By_{2t} = \beta' \varepsilon_t. \tag{1a'}$$

Now, if we naively fit the level-VAR model in Equation 1, we will treat that system as a reduced form and assume that y_{t-1} is predetermined. In fact, the y_{t-1} is endogenous; it embodies the endogeneity evident in Equation 1a'. This can be seen by adding and subtracting $B\varepsilon_{2t}$ to the right side of (1a):

$$y_{1t} = (By_{2t-1} + B\varepsilon_{2t}) - (B\varepsilon_{2t} - \varepsilon_{1t}). \tag{1c}$$

Note from Equation 1b that the first two terms on the right side of this equation are equivalent to By_{2t}. So Equation 1c implies

$$y_{1t} = By_{2t} - B\varepsilon_{2t} + \varepsilon_{1t} = By_{2t} + \beta' \varepsilon_t. \tag{1d}$$

Thus we have shown that the endogeneity that is present in Equation 1a' also is present in Equation 1a.

This endogeneity has seemingly serious consequences for causal inference. y_{2t-1} satisfies the orthogonality condition for a predetermined variable, namely, $E(\varepsilon_{1t}y'_{2t-1}) = 0$. But since y_{2t-1} is nonstationary, the sample covariance $T^{-1}\sum_1^T \varepsilon_{1t}y'_{2t-1}$ does not converge to zero. Rather

$$T^{-1}\sum_1^T \varepsilon_{1t}y'_{2t-1} \xrightarrow{d} \int_0^1 dB_1 B'_2 \tag{2}$$

where $\rightarrow d$ means converges in distribution, $B_1(r \times 1)$ and $B_2(n\text{-}r \times 1)$ are subvectors of Brownian motion $B = (B_1', B_2')' \sim BM(\Sigma_{\varepsilon\varepsilon})$, B_1 is the limit process of partial sums of ε_{1t}, and B_2 is the limit process of $T^{-1/2}y_{2t-1}$ (Phillips 1988a).[6] If the correlation between ε_{1t} and ε_{2t} is nonzero so that $\Sigma_{\varepsilon\varepsilon}$ is not block diagonal, the limit processes B_1 and B_2 will be correlated Brownian motions. This is due to the effect in the limit of the "endogeneity" of the regressor y_{2t-1} in Equation 1a.

Consider the OLS estimation of the B matrix in that equation. According to asymptotic distribution theory (proposition 18.1 in Hamilton 1994, 547–8), the limit distribution for the respective coefficients will be given by

$$\begin{aligned} T(\hat{B} - B) &= \Big(T^{-1}\sum_{1}^{T} \varepsilon_{1t} y_{2t-1}'\Big)\Big(T^{-2}\sum_{1}^{T} y_{2t-1} y_{2t-1}'\Big)^{-1} \\ &\xrightarrow{d} \left(\int_0^1 dB_1 B_2'\right)\left(\int_0^1 B_2 B_2'\right)^{-1} \end{aligned} \tag{3}$$

The right side of Equation 3 can be decomposed into

$$\left(\int_0^1 dB_{1\cdot 2} B_2'\right)\left(\int_0^1 B_2 B_2'\right)^{-1} + \sum_{12} \sum_{22}^{-1} \left(\int_0^1 dB_2 B_2'\right)\left(\int_0^1 B_2 B_2'\right)^{-1} \tag{4}$$

where $B_{1\cdot 2} = B_1 - \Sigma_{12}\Sigma_{22}^{-1}B_2 = BM(\Sigma_{11\cdot 2})$ with $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$. The first term in the expression is the limit distribution of the optimal estimator of B under Gaussian errors in Equation 1. The second term is the "simultaneous equation bias" that is due to the endogeneity of the nonsta-

[6]Brownian motion—also known as a "Weiner process"—is the continuous time version of the discrete time random walk. It has three main features: the Markov property (the probability distribution for all future values of the process depends only on its current value and is unaffected by past values of the process), independent increments, and changes that are normally distributed. Formally, if z(t) is Brownian motion, then any change in z, Δz, corresponding to time interval Δt, is such that

$$\Delta z = \varepsilon_t \sqrt{t}$$

where ε_t is N(0,1) and $E(\varepsilon_t \varepsilon_s) = 0$ for $t \neq s$. These conditions imply that the variance of the change in Brownian motion grows linearly with the time horizon just as it does in a discrete-time random walk. The continuous time representation is obtained by letting Δt become small:

$$dz = \varepsilon_t \sqrt{dt};$$

and, $E(dz) = 0$, and $\text{Variance}(dz) = E[(dz)^2] = dt$. The concept of vector Brownian motion is just an extension of these ideas that includes provision for the covariance between the respective terms. The key role of stochastic integration in relation to the derivation of the relevant limit distributions here is explained in such works as Phillips 1988b. The present explanation is condensed from Dixit and Pindyck 1994 (Chapter 3). See also Hamilton 1994, 474–9, Granato and Smith 1994 (fn. 4), and Malliaris and Brock 1982: 36–8.

tionary regressor y_{2t-1} in Equation 1a. It is this term that causes the limit distribution for $\hat{\beta}$ to be miscentered and skewed and to depend on nuisance parameters. Again, the nonstandard limit distribution in Equation 3 means that standard t and F statistics cannot be used to make inferences about the statistical significance of $\hat{\beta}$.

1.2 Discussion

VAR was pioneered in the social sciences by Sims (1980). The method always has baffled mainstream time series analysts because, as Harvey (1990, 83) puts it, "The usual approach adopted by VAR aficionados is . . . to work in levels, even if some of the series are nonstationary." Hence, by implication, VAR modelers are prone to the problem outlined above. But how serious is this problem? For example, are all the Granger causality tests that VAR modelers use necessarily in error?

To begin with, note that in all cases working in levels provides estimates that are asymptotically efficient, a result that owes to Fuller (1976, 347–9) in the univariate case. And, as long as the order of the VAR is greater than one, all estimated coefficients are asymptotically normal, and all t-tests of hypotheses about individual coefficients are asymptotically valid (Hamilton 1994, 553). Sims, in advocating level VARs, clearly believed Fuller's result held in the generalized VAR case. He also clearly believed that typical transformations of variables are merely convenient simplifications that potentially produce losses of information. In both cases, he was correct. Sims, Stock, and Watson (1990) show that Fuller's results generalize to a VAR without drift; they extend the results with similar findings to the case with drift. Seen in this light, differencing is merely a convenience for making sure that standard distribution theory applies in all situations.[7]

Sims' intuition regarding information losses resulting from differencing also turned out to be correct, however. Differencing rids series of a long-term component and, therefore, makes it impossible to model the respective elements of time series. In this sense, a VAR in levels is more consistent with concepts like cointegration, the idea that two or more time series share a common trend in levels (Hamilton 1994, 579–80).

In addition, a VAR in levels has a useful property that plagues the conventional cointegration and error correction approaches to model building. Cointegration and unit root tests provide knife-edge results insofar as they lead to yes or no decisions regarding the trend properties of a (set of) time series. A level VAR user need not make these knife-edge decisions. For example, the coefficient on the first lag of the lhs variable is allowed to take on

[7]Differencing may increase the speed of convergence and thus improve small sample performance (Hamilton 1994, Chapter 18).

an estimated value without an arbitrary judgment that the series is I(1) or I(0). Furthermore, it is not hard to conceptualize a system with nearly cointegrating relationships among variables for which the VAR in levels does not make such arbitrary restrictions. For this system, the moving average representation (MAR) for the level VAR will chart responses that may be somewhere in between long-term and short-term dynamics. It is conventional wisdom among VAR analysts that systems with unit roots often exhibit covariance stationarity in precisely these cases with the MAR providing dynamic responses in a number of situations with dampening rather than explosive dynamics.

Once more, the primary cost in modeling a VAR in levels is that even if its order is greater than one, tests of a linear combination of coefficients—some of which are coefficients on nonstationary variables—may have statistics with nonstandard limit distributions. To see this more clearly, consider the following representation:

$$y_t = \alpha + \sum_{s=1}^{S} \beta_s y_{t-s} + e_t \tag{5}$$

where y and e again are n-vectors, β_s now is an n by n coefficient matrix for each lag s, and α is a n-vector of constants. Suppose that instead of this representation we differenced all the variables in y_t and reestimated the system of equations. This would effectively force the $\beta_{s=1}$ coefficients on the lagged y's in Equation 5 to equal unity. The differenced representation also would require the coefficients for lags two through *S* to be linear transformations of the original elements of $\beta_{s>=1}$ while again stipulating that the coefficients on the lagged lhs variables sum to unity.

With these facts in mind, we can write the level VAR in Equation 5 in the following way:

$$y_t = \alpha + \rho y_{t-1} + \sum_{s=1}^{S-1} b_s \Delta y_{t-s} + \mu_t \tag{6}$$

where ρ is an n by n matrix. In keeping with the derivation above, Sims, Stock, and Watson (1990) show that in terms of the representation in Equation 6, the $b_{s>=1}$ will have the usual Gaussian limit distributions (assuming that the roots of the corresponding characteristic equation lie outside the unit circle) but the limit distributions for the ρ will be nonstandard. Moreover, the coefficients in $\beta_{s>=1}$ are linear transformations of coefficients in $b_{s>=1}$ meaning that the VAR in levels provides coefficients in $\beta_{s>=1}$ that follow standard Gaussian distributions (see Hamilton 1994, 550–4).

As regards hypothesis testing, the implication is that Wald tests on $b_{s>=1}$ or $\beta_{s>=1}$ will have standard distributions. Thus model order can be tested with conventional statistics because these statistics rarely if ever involve restrictions on ρ. A test for a VAR(S) against a VAR(S+1) with S>1 in Equation 6 will have a limiting distribution that is standard.

The difficulty will arise in Granger causality tests because they involve restrictions based *in part* on the coefficients on the first variable on the rhs of Equation 6, ρy_{t-1}. The statistics for these tests will have nonstandard limit distributions; these distributions will have fatter tails than the standard ones. What is even more vexing is the fact that these nonstandard limit distributions may depend on nuisance parameters in the level VARs.[8]

2. Solutions

Several approaches to estimating and testing hypotheses with VARs have been suggested.[9] One of these was proposed by Sims, Stock, and Watson (1990). The intuition behind the Sims, Stock, and Watson method is straightforward, if the actual calculations are not. A Granger causality test implies restrictions, under the null of noncausality, on ρ and b in Equation 6. To test, say, that variable y_{2t} Granger causes y_{1t}, requires that the null be set up such that a single element from ρ and S-1 elements of $b_{S>=1}$ be set equal to zero. Thus, standard distribution theory applies to the S-1 elements; in particular, these elements are distributed as an F. The inclusion of the ρ coefficient, on the other hand, will necessitate the use of a nonstandard distribution. To test the null hypothesis of non-Granger causality a composite distribution must be constructed, one that is made up both of the nonstandard and standard distributions just mentioned; finding the p-values associated with such tests requires integrating over the combination of the two distributions. In practice, these integrations are usually intractable and, therefore, must be carried out using numerical methods such as Gaussian quadrature or, more generally, Monte Carlo integration.[10]

[8]Sims, Stock, and Watson (1990, 129 and following) study a trivariate VAR in which each variable has a unit root with nonzero drift in its univariate representation. They consider four cases: (i) no cointegration, time trend excluded from model; (ii) no cointegration, time trend included; (iii) cointegration, time trend excluded; and (iv) cointegration, time trend included.

[9]The only two articles we have found that propose methods to deal explicitly with level VARs are Sims, Stock, and Watson (1990) and Phillips (1995).

[10]For instance, Sims, Stock, and Watson (1990) show that for their case iii (see our footnote 8), if there are two cointegrating vectors in the 3 X 3 system, all coefficient estimators will be asymptotically normally distributed and all test statistics will have the usual asymptotic chi-square distributions. The implication is that a Granger causality test for this case will have an asymptotic χ_p^2/p distribution, where p is the order (lag length) of the matrix polynomial that defines the VAR model. This is true as long as there is a particular cointegrating vector involving a particular rhs variable. If no such cointegrating vector exists, the test statistic will have a nonstandard distribution (1990, 132–6, fn. 3: 136). The authors also provide the formula for calculating the composite F statistic for one set of conditions related to case iii (136). A more general analysis of this kind can be found in Toda and Phillips 1993.

In their article "Interpreting the Evidence on Money Income Causality," Stock and Watson (1989) illustrate this approach. First, the data are pretested for unit roots, deterministic trends, and common trends. Next, on the basis of the results of step one, an appropriate composite distribution is identified, and numerical integration is used to calculate the corrected p values for certain individual test statistics. In this way, the confusion about the relationship between money and income is clarified; conflicting results are shown to be the result of the application of the incorrect distribution theory in cases where pretesting and appropriate adjustments in p values were not made in existing works.[11]

While writers like Hamilton maintain the Sims, Stock, and Watson approach is promising (1994, 653), it has a number of drawbacks. First, this approach is subject to pretest bias (1990, 136–7; Stock and Watson 1989, 178; Phillips 1995, 1028–9). Discovering the trend properties in step one is prone to error which, in turn, can lead to misapplication of (non)standard distribution theory.[12] Second, determining when nonstandard distribution theory is required is difficult. Sims, Stock, and Watson derived results only for a trivariate system and then for a small number of cases associated with that system. As Phillips (1995, 1053) notes, the Sims, Stock, and Watson analysis does not "provide an asymptotic theory that justifies the general use of VAR regressions for causality testing at least in correctly specified models."[13]

A second approach is Johansen's maximum-likelihood method, or what is sometimes called "reduced-rank regression." To implement this approach, one begins with the assumption that each of the individual y_{it} in Equation 1 are I(1). Then, as with the Stock and Watson approach, one proceeds to use pretests to determine the number, h, of cointegrating relationships between

[11]More specifically, Stock and Watson (1988) show that the confusion about money-income causality is a result of scholars failing to take into account the use of differences and linear and polynomial trends in the presence of unit roots and cointegrated series, or failing to note the need for the use of nonstandard distribution theory in some specifications but not in others. They explain how Monte Carlo integration is used to calculate the nonstandard p values in the appendix to their paper.

[12]Consider, for instance, testing for deterministic trends. This is straightforward insofar as standard distribution theory applies to the coefficients on the respective variables (Hamilton, 1994, Chapter 16). The inclusion of time trends, however, will influence the nonstandard distributions of other hypothesis tests when other coefficients have nonstandard distributions; the existence of deterministic trends can change the results of standard (stationary) and nonstandard Granger causality tests if these trends are ignored. Thus, specifying the exact nature of these "nuisance" parameters is crucial. Unfortunately, there is no clear and best way to determine whether a time series has a deterministic vs. stochastic trend. The respective tests have low power, and, in small samples such tests are very fragile (Dejong et al. 1992). A good introduction to the problem of testing for unit roots are the papers by Evans and Savin (1981, 1984).

[13]Stock and Watson kindly supplied us with the code they used for their study of money and income. However, that code applied only to the specific system that they had constructed on the basis of their pretesting of the relevant money-income series.

the elements of y_t. It is well known that if the system involves h cointegrating relationships, then the coefficient ρ on y_{t-1} in Equation 1 can be expressed as $\rho = A^*B$, where A and B are n by h and h by n matrices, respectively. Johansen then shows how to obtain the maximum likelihood estimates of the parameters in Equation 1 subject to the constraint that $\rho = A^*B$ (and that the distribution of the system's errors is Gaussian). Juselius and Hargreaves (1992) and others use this method to study the long-run properties of money demand and other economic processes. On the basis of the sequence of hypothesis tests just described, they characterize the long-run relations between certain monetary and income aggregates.[14]

There are several problems with the Johansen method. Most important is that of correctly identifying the number of cointegrating vectors in the system. If this is not done, the maximum-likelihood estimates (and, therefore, the hypothesis tests) can be seriously in error; the results of any hypothesis test about causal relations in the data can be inaccurate if one errors in determining the dimensionality of the cointegration space. The severity of the problem hinges on such things as the relative variances associated with the different cointegration vectors. There are other problems with the Johansen method: for example, apparent sensitivity to the lag-length specification in the error-correction setup.[15]

[14]This description is based on Juselius and Hargreaves (1992) and Hargreaves (1994). The former source explains the connection between Johansen's method and the probability approach of Hendry and Richard (1983); the latter explains the relation between Johansen's method and the Engle and Yoo three-step method (1991). Suppose z_t is a vector time series with dimension p. Assume the lag order is 2 so the process is simply:

$$z_t = A_1 z_{t-1} + A_2 z_{t-2} + e_t \qquad e_t \sim NID_p(0,\Sigma)$$

Then the Johansen method focuses on the error-correction form:

$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + \Pi z_{t-1} + e_t$$
$$\Delta z_t = \Gamma_1 \Delta z_{t-1} + \alpha\beta' z_{t-1} + e_t$$

where $\gamma_1 = -A_2$ and $\alpha\beta'$ is the matrix product discussed in the text (and equal to $\pi = 1 - A_1 - A_2$). Juselius and Hargreaves (1992) show how dummy variables, deterministic time trends, and more complex error processes are incorporated into this setup. They also describe the likelihood function, which is made up of product-moment matrices which, in turn, are composed of the levels residuals and difference residuals from the auxiliary regressions. The same source describes the different hypothesis tests which are performed with the Johansen method and the test's connection to problems like overparameterization. A useful discussion of the general problems in estimating cointegrating relationships is Saikkonen (1991).

[15]An evaluation of Johansen's approach can be found in works such as Hargreaves (1994). For instance, after comparing the performance of it and various single-equation approaches, Hargreaves concludes, "Results can change dramatically if one changes the dimensionality of the cointegration space" (1994, 103). In a companion piece he and Juselius add, "Since all subsequent [hypothesis] tests are valid under the condition that the [cointegration] rank is correctly determined, the choice of r is very crucial in the analysis" (1992, 268; see also 271 and their footnotes 2 and 3).

Phillips and his associates have proposed an alternative approach. Originally, this method—FM-OLS—was designed to estimate cointegrating relationships in single equations. Recently the approach has been extended to cover the more general time series situation in which a VAR might contain unit roots and(or) cointegrating vectors. This is the new FM-VAR estimator (Phillips 1995). Because the older FM-OLS is based on the same logic and is much easier to explain than FM-VAR, we describe it in detail. We then sketch the properties of the FM-VAR estimator.

Consider the following model[16]:

$$y_t = \beta x_t + u_{1t} \tag{7}$$
$$x_t = x_{t-1} + u_{2t}$$

where y_t and x_t are vectors for which the regressor variables are not cointegrated, and their differences are stationary (u_{1t} and u_{2t} are stationary processes and u_{2t} is white noise so x_t is a random walk). Say that the model describes a system of relationships, and y is dimension r, and x is dimension p-r. Then the model can be rewritten as a simultaneous equation system:

$$\begin{bmatrix} A_1 & A_2 \\ B_1 & B_2 \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix}_{p-r}^{r} = \begin{bmatrix} I & A_1^{-1}A_2 \\ 0 & I \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix}_{p-r}^{r} = \begin{bmatrix} & u_{1t} \\ \Delta^{-1} & u_{2t} \end{bmatrix}_{p-r}^{r} \tag{8}$$

where Δ is the difference operator. Let u_{1t} be AR(1):

$$u_{1t} = \rho u_{1t-1} + e_t \qquad |\rho| < 1 \tag{9}$$

and suppose that

$$\begin{bmatrix} e \\ u_{2t} \end{bmatrix} \sim NID\left[\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \varphi\sigma_1\sigma_2 \\ \varphi\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}\right]$$

so e_t and u_{2t} are both stationary. Therefore, u_{1t} is also stationary. Since its first difference is stationary, x_t is I(1). And, because y_t is a linear combination of an I(1) variable and an I(0) variable, it also is I(1). Since the linear combination of these two vectors is I(0), they are cointegrated.[17] As we saw earlier, the asymptotic distribution of the OLS estimator of β becomes

[16]The following is a condensation of Hargreaves 1994, Section 2.3.

[17]The equation system in Equation 8 is equivalent to that in Equation 7 with $\beta = -A_1^{-1}A_2$. The linear combination of y_t and x_t is u_{1t}, which is I(0). Hence the two vectors are cointegrated by construction. It should be noted that the model in Equation 7 is not assumed to be the structure of economic reality but rather an approximation of the same (". . . in a similar way to a Box-Jenkins time series ARIMA model paralleling a causal structural model. One assumes that the effects of other stationary variables are modeled equivalently by the ARMA process on the errors . . . [in this sense the model] is a device used to efficiently estimate the cointegrating relationship" Hargreaves 1994, 95.)

$$T(\hat{\beta}-\beta)\underset{\rightarrow}{d}\left(\int B_2^2\right)^{-1}\frac{\sigma_1\sqrt{1-\varphi^2}}{1-\rho}\int B_2 dW+\frac{\varphi\sigma_1}{(1-\rho)\sigma_2}\int B_2 dB_2+\frac{\varphi\sigma_1\sigma_2}{1-\rho} \qquad (10)$$

where W = BM(I) is independent of B_2. The last two terms again are due to serial correlation in the unit root regressor and endogeneity. It is these two terms that FM-OLS tries to estimate and asymptotically remove.[18]

This is done in the following way. First, we stack the errors into one vector

$$u_t=\begin{bmatrix}u_{1t}\\ u_{2t}\end{bmatrix}_{p-r}^{r} \qquad (11)$$

We then form the cumulative sum of u_t which is a multivariate random walk. In terms of continuous time we have

$$w_T(\tau)=\frac{1}{T}\sum_{t=1}^{[T_\tau]}u_t\underset{\rightarrow}{d}B(\tau)=\begin{bmatrix}B_1\\ B_2\end{bmatrix}_{p-r}^{r}=BM(\Omega) \qquad (12)$$

for $\tau\in[0,1]$ and $T\rightarrow\infty$. The covariance matrix, Ω, of the Brownian motion is called the "long-run covariance matrix." It is defined as

$$\Omega=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}E(u_j u_t'); \qquad (13)$$

as such it represents the sum of all the covariances backwards and forwards of u_j and u_t. Ω can be decomposed into a contemporaneous variance and sums of autocovariances

$$\begin{aligned}\Omega&=\lim_{T\to\infty}\frac{1}{T}\left\{\sum_{t=1}^{T}E(u_t u_t')+\sum_{t=2}^{T}\sum_{j=1}^{t-1}E(u_j u_t')+\sum_{t=2}^{T}\sum_{j=1}^{t-1}E(u_1 u_j')\right\}\\ \Omega&=E(u_o u_o')+\sum_{t=1}^{\infty}E(u_o u_t')+\sum_{t=1}^{\infty}E(u_t u_o') \qquad (14)\\ \Omega&=\sum+\Gamma+\Gamma'\end{aligned}$$

[18]The first term in Equation 10 is independent of the other two since B_2 and W are independent. If x_t is strictly exogenous φ is zero and the second and third terms in Equation 10 are both zero.

where Σ is the contemporaneous covariance matrix. Define

$$\Delta = \Sigma + \Gamma \tag{15}$$

Then we can write Ω and Δ conformably with u_t:

$$\Omega = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{bmatrix}_{p-r}^{r} \text{ and } \Delta = \begin{bmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{bmatrix}_{p-r}^{r} \tag{16}$$

Given what we have assumed about the structure of the errors we have:

$$\Sigma = \begin{bmatrix} \sigma_1^2/(1-\rho^2) & \varphi\sigma_1\sigma_2 \\ \varphi\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix} \tag{17}$$

$$\Omega = \begin{bmatrix} \sigma_1^2/(1-\rho^2) & \varphi\sigma_1\sigma_2/(1-\rho) \\ \varphi\sigma_1\sigma_2/(1-\rho) & \sigma_2^2 \end{bmatrix} \tag{18}$$

$$\Gamma = \begin{bmatrix} \rho\sigma_1^2/\left[(1-\rho)(1-\rho^2)\right] & 0 \\ \rho\varphi\sigma_1\sigma_2/(1-\rho) & 0 \end{bmatrix} \tag{19}$$

Define the long-run variance of u_{1t} conditional on u_{2t} as

$$\Omega_{1\cdot 2} = \Omega_{11} - \Omega_{12}\Omega_{22}^{-1}\Omega_{21} \tag{20}$$

and the bias due to the endogeneity of the regressors as

$$\Delta_{21}^{+} = \Delta_{11} - \Delta_{22}\Omega_{22}^{-1}\Omega_{21} \tag{21}$$

Because it is second-order stationary, the long-run autocovariances are y_t's spectral density evaluated at the origin. And a kernel estimator with an appropriate bandwidth specification can be employed along with the OLS estimates of the errors to make the indicated calculations (see Lee and Phillips 1994). Denoting these estimated autocovariances with a hat, these calculations yield

$$\hat{\Omega}_{1\cdot 2} = \hat{\Omega}_{11} - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21} \tag{22}$$

and

$$\hat{\Delta}_{21}^{+} = \hat{\Delta}_{11} - \hat{\Delta}_{22}\hat{\Omega}_{22}^{-1}\hat{\Omega}_{21}. \tag{23}$$

The original dependent variable then is transformed to be

$$y_t^+ = y_t - \hat{\Omega}_{12}\hat{\Omega}_{22}^{-1}\hat{u}_{2t}. \qquad (24)$$

The FM-OLS estimator, then, is:

$$\hat{\beta}^{fm} = \left\{\sum_{t=1}^{n}(y_t^+ x_t' - (0\hat{\Delta}_{21}^{+\prime}))\right\}\left(\sum_{t=1}^{n} x_t x_t'\right)^{-1} \qquad (25)$$

With this method, one can obtain asymptotic standard normal t statistics which can be used for hypothesis tests which include the coefficients on nonstationary variables. However, the distribution of the estimator is non-normal and skewed. Hence confidence intervals cannot be calculated in the conventional way.[19]

The Fully-Modified estimator is very complicated. It involves transforming the standard VAR model into an equivalent form on which one focuses for estimation and hypothesis tests. Say the level VAR model is expressed as

$$y_t = J(L)y_{t-1} + \varepsilon_t \qquad (26)$$

where $J(L) = \sum_{i=1}^{k} J_i L^{i-1}$ and L the lag operator. We analyze the equivalent form:

$$y_t = J^*(L)\Delta y_{t-1} + Ay_{t-1} + \varepsilon_1 \qquad (27)$$

where $J^*(L) = \sum_{i=1}^{k-1} J_i^* L^{i-1}$ and $J_i^* = -\sum_{h=i+1}^{k} J_h$, $A = J(1)$ and Δ is the difference operator.[20] This model is transformed into another (orthogonalized) coordinate system, and the FM-VAR estimator and its limit theory is derived. Then these results are translated back into parallel results for the original coordinate system corresponding to the form in Equation 27. The

[19]Since it is conditional on $\sigma(B_2(\gamma), 0 < \gamma \leq 1)$ the estimator is asymptotically normally distributed. Unconditionally it is a mixture of normals like all so-called "unit root estimators." Phillips (1995) derives the limit theory for FM-OLS, relates this theory to Wald tests, and shows how deterministic regressors can be incorporated in the estimation.

[20]Note that Equation 27 has the equivalent VECM representation: $\Delta y_t = J^*(L)\Delta y_{t-1} + (A\text{-}I)y_{t-1} + \varepsilon_t$ where the notation is that used in the text.

actual estimator of the coefficients β is a more complex, alternative matrix product than its OLS counterpart; it amounts to a transformation of the original lhs variables in order to remove the endogeneity bias resulting from unit roots and cointegrating vectors in the original VAR system. The FM-VAR estimator in the terms of the original coordinates is:

$$\hat{F}+ = \left[Y'Z \vdots Y'Y_{-1} - \hat{\Omega}_{\varepsilon y}\hat{\Omega}_{yy}^{-1}(\Delta Y'_{-1}Y_{-1} - T\hat{\Delta}_{\Delta y \Delta y})\right](X'X)^{-1} \tag{28}$$

where the $\vdots$ represents a partitioned matrix, Z is a vector of first differences in lags of y_{t-1}, and the $\hat{\Omega}$ and $\hat{\Delta}$ are kernel estimates of the appropriate long run covariance matrices akin to those described above for FM-OLS. In this way, one transforms the original data in terms of corresponding auto-covariance matrices to obtain a new estimate of the coefficients in Equation 27 in a manner parallel to that which was done by the FM-OLS estimator (cf. Eqns. 28 and 25).

While the theory behind Fully-Modified estimation is somewhat complicated, the intuition that underlies it is straightforward. As discussed earlier, when nonstationary variables are included in Equation 5, standard OLS will estimate β_1 with second-order bias. In particular, its limiting distribution will be asymetric and leptokurtic. Loosely speaking, this bias is caused by endogeneities in the predetermined variables resulting from those variables' nonstationarity. To see this, note that in Equation 5, by construction, plim $T^{-1}\Sigma y_{t-1}e_t$ does not converge to zero. The long-run covariance between the regressors and the errors is not zero. It is this nonzero long-run covariance matrix that leads to the biased limiting distributions of the coefficient estimates. FM-VAR, then, deals with the fact that the VAR is not a 'reduced form' by adopting a semiparametric estimation technique that corrects the OLS-VAR formula by accounting for the endogeneity bias. Importantly, these adjustments are made without prior information about the cointegrating space, thus obviating the need for pretests and the potential for pretest bias.

Phillips explains that the FM-VAR estimator is preferable to its FM-OLS counterpart. This is because (1) in nonstationary directions the latter contains second-order bias (that is, simultaneous-equation bias), and (2) it entails a composite of a matrix unit root distribution and a mixed normal in the estimation of the system's unit roots (1995, Remark 5.8). He also notes that when the original VAR contains stationary components, the FM-VAR estimates of the respective coefficients will have the same asymptotic distribution as the level VAR OLS estimates. Finally, Phillips shows that the Wald test on FM-VAR coefficients has a limit distribution that is a linear combination of chi-square variates. This limit variate is bounded above by a χ^2 statistic with degrees of freedom equal to the number of restrictions being tested. Therefore, conservative but asymptotically valid hypothesis tests like those associated with the concept of Granger causality can be performed with the FM-VAR approach.[21]

Practically speaking then, this third approach involves three steps. First, one rewrites his or her model in a form similar to Equation 27. Standard lag length tests then are implemented along the lines suggested earlier. Next FM-VAR is used to obtain estimates of the coefficients in the transformed model; these estimates are based on those of the long-run autocovariances described above. Finally, Granger causality tests are performed on fitted models; the joint statistical significance of the appropriate FM-VAR estimated coefficients is assessed. This yields a test statistic for the null of non-Granger causality whose distribution is bounded above by the distribution of a chi-square random variable with degrees of freedom equal to the number of imposed linear restrictions. In other words, one may use the test statistic and the appropriate chi-square distribution to determine a p value for the null of non-Granger causality that is greater than or equal to the p value we would obtain if we were to use the test statistics' true (but complicated) distribution.

FM-VAR has a number of advantages over the Sims-Stock-Watson and Johansen approaches. Most important, FM-VAR avoids pretest bias while at the same time achieving equivalencies with the Johansen method:

> In VAR models with some unit roots and cointegrated variables (a composite system) the FM-VAR estimates of the identified components of the cointegrating matrix have a mixed normal limit theory which is equivalent to the optimal estimator in Phillips (1991a) or the reduced rank regression estimator in Johansen (1988). Moreover, optimal estimation of the cointegration space is attained in FM-VAR regression VAR regression *without knowledge* of the dimension of the cointegration space and *without pretesting* for the number of cointegrating vectors. Thus an investigator can perform unrestricted regression by FM-VAR and effectively disregard the I(1) and I(0) nature of the data. Any cointegrating relations are implicitly estimated as if one was performing a maximum likelihood estimation of the model with the cointegrating rank known correctly. (Phillips 1995, 1056, emphasis in original)

Second, in comparison to the Sims, Stock, and Watson approach, the need to derive an expression for the exact nonstandard limit distribution of the

[21]The other main properties of the FM-VAR estimator are: (1) when there is cointegration in the system its limit theory is normal (asymptotically equivalent to OLS) for stationary coefficients and mixed normal for *all* of the nonstationary coefficients including unit roots (there are no unit root distributions, and there is no asymptotic bias in the estimation of the cointegration space in the FM-VAR limit theory) and (2), when the system has a full set of unit roots, the FM-VAR estimator of the complete unit root matrix is hyperconsistent in the sense that the rate of convergence of the estimator exceeds the O(T) rate of the OLS and MLE estimator (Phillips 1995, 1025–6). As regards its relative virtues in relation to FM-OLS, Phillips notes that the latter is applicable in models with either full rank or cointegrated I(1) regressors and in models with stationary regressors (1056).

estimator and test statistics is lessened to some degree. A Granger causal hypothesis involving stationary and nonstationary coefficients in a FM-VAR estimated system produces a Wald statistic with a limit distribution which is a linear combination of independent chi-square variates. The χ_q^2 distribution (for a test of q total restrictions) is an upper bound for this statistic. Hence it can be used to make conservative evaluations of the hypotheses of non-Granger causality (Philips 1995, Section 7, especially 1057). Moreover, since the FM-VAR estimator is unaffected by their presence in the equation system, the causality tests need not make provisions for drifts, trends, and other nuisance parameters. With FM-VAR, then, we may not obtain exact p values for hypothesis tests, but we can make some useful inferences about the validity of our causal claims.[22] Finally, Phillips contends that because it does not employ pretests and sequential testing, FM-VAR is more "in the spirit" of VAR than the Johansen method (1995, 1029).

There are several issues surrounding FM-VAR. To begin with, kernel estimation of the spectral density associated with the long-run covariances is required. Often, the Parzen kernel along with an automatic bandwidth selector is used for this purpose. Some researchers suggest using alternative kernel estimators and additional procedures like VAR prewhitening. Most of these recommendations are based on comparisons of FM-OLS estimators rather than FM-VAR estimators, however. But work on the properties of the FM estimators really has just begun.[23] In addition, one must exercise care in

[22]Of course, in principle, with the Sims, Stock, and Watson approach we can obtain an *exact* p value for our Granger causality test statistics. This p value would be arbitrarily accurate, although it would still be subject to specification error due to the poor performance of pretest estimators and other problems.

[23]The long-run covariance concept is a multivariate extension of the same idea that is at the heart of the KPSS test for stationarity and the modified R/S statistic and variance ratio tests for fractional integration. (On this connection see Box-Steffensmeier and Smith 1996.)

Because the long-run covariances are infinite sums we can only approximate their values. Since the relevant series is second-order stationary, its long run covariances are equivalent to their spectral densities evaluated at the origin times a constant (Lee and Phillips 1994). Kernel estimation of these spectral densities is possible; the bandwidth for this estimation normally is chosen by an automatic procedures (see, for instance, Hamilton 1994 (Chapter 6, especially 165 and following); see also Ng and Perron 1995). Below we show that the use of alternative procedures for bandwidth selection does not appreciably affect the performance of the FM-VAR estimator.

Illustrative of the alternative procedures which have been suggested in FM-OLS estimation is that of Cappuccio and Lubian (1995). They propose capturing some of the dependence in the $\hat{u}_t = \hat{u}(\theta)_t$ which are obtained from a consistent estimate of θ in the class of kernel consistent estimators of Ω and Γ. In particular, Cappuccio and Lubian propose to fit a VAR(1) model of the u and use the resulting residuals to obtain the long-run covariances matrixes. The VAR(1) residuals presumably are closer to white noise process than $\hat{u}_t$. In a Monte Carlo experiment Cappuccio and Lubian find that VAR prewhitening of FM-OLS yields empirical t-ratios that are better approximations to the standard normal t distribution than those obtained with conventional methods. For more details about kernel estimation procedures and a brief review of some new works evaluating FM estimation, see in addition to Cappuccio and Lubian (1995), Phillips (1995, 1025). See also Hargreaves (1994.)

interpreting and using the Granger causality test statistic produced by FM-VAR. Again, the respective p values are calculated using a chi-square distribution that is an upper bound, not the true distribution of the test statistic. The actual p value for a Granger causality test is less than that indicated by the relevant χ^2 distribution. This means, in effect, that FM-VAR Granger causality tests are biased toward confirming the null of non Granger causality; *ceteris paribus*, users of FM-VAR will tend to accept the hypothesis of no Granger causal relationship. Whether this is a serious problem—whether FM-VAR based Granger causality tests lack statistical power in finite samples—remains to be determined in Monte Carlo experiments like those we conduct below (cf. Phillips 1995, 1055).[24,25] Finally, the general pitfalls of Granger causality testing—omitted variables bias, mistaken inference due to sampling design, etc.—all still apply (Freeman 1983).

3. Implications for Political Science

Political science certainly appears to be prone to the problems outlined above. Many of us have not addressed the possibility that our political time series contain unit roots or are cointegrated. We also have failed to make any provisions for nonstationary coefficients in our level VARs and, concomitantly, for nuisance parameters in our estimator's limit theory. In our desire to demonstrate the robustness of our results to a variety of specifications, we have differenced variables and included various kinds of trend terms in our level VARs without taking into account how the limit theory for test statistics varies for alternative specifications.[26] Critics like Ostrom and Smith (1993) and Granato and Smith (1994), therefore, are right to question the validity of our findings. Where there are sound theoretical and conceptual

[24]The use of such bounds in statistical inference is not unusual. One encounters them in time series regression analysis with non-Gaussian errors, for instance (Hamilton 1994, 214).

[25]Kitamura (1994) constructs a test statistic for FM-VAR estimators and proves it has a simple distribution whenever the process under consideration is stationary or cointegrated (if the process is full-rank integrated the distribution of the statistic is again bounded by a chi-square distribution). Using his method, therefore, in integrated systems, it is possible to obtain *exact* p values when conducting Granger causality tests. Kitamura's approach is still under development; it is not yet available for use in investigations like the present one.

[26]Political science assessments of trend properties of time series usually are incomplete and unsystematic. Some of our studies use level data and allow for the existence of deterministic linear trends (Freeman, Williams, and Lin 1989); other investigations test for unit roots and for the existence of common trends (Ostrom and Smith 1993; Durr 1993a); still others use level data and test neither for deterministic trends nor univariate or multivariate stochastic trends (Williams 1990). The tests which are done usually are quite unreflective. For instance, in testing for deterministic trends, constant rates of change are assumed implicitly. The possibility of polynomial trends or that long-term rates of change, for example, in growth rates, are deterministic but non constant also often are ignored. In general, while stochastic trends are increasingly understood and evaluated, deterministic trends are often ignored and no meaningful substantive interpretation of them is provided. For a further elaboration of this critique see Freeman and Stimson (1994).

reasons to expect cointegration and (or) unit roots, our causality tests may well have been plagued by the biases described above. With FM-VAR, we now can gauge the accuracy of these causal inferences. It is to this vitally important task that we now turn.

4. Evaluation and Reanalysis

4.1 Monte Carlo Experiment

In order to better explain the nature of FM-VAR and to gauge its finite sample performance, we begin with a brief report of results from several Monte Carlo experiments. We generated data for these experiments using the simple two equation system:

$$y_{1t} = a_1 y_{1,t-1} + \varepsilon_{1t} \tag{29.1}$$

$$y_{2t} = a_2 y_{2,t-1} + b y_{1,t-1} + \varepsilon_{2t} \tag{29.2}$$

The errors were specified to have unit variance and to be serially and contemporaneously uncorrelated. We chose the parameter values to examine four important types of stochastic processes: full-rank integrated, nearly integrated, stationary, and mixed stationary-nonstationary. The particular parameter values used for our experiments will be discussed in detail below. Samples of 25, 50, 100, and 500 were studied for each type of system.

Each data set was used to estimate a level, second-order VAR that included neither constants nor deterministic trends. In particular, we analyzed the following system:

$$[y_{1t}\ \ y_{2t}]' = \theta_1\,[y_{1t-1}\ \ y_{2t-1}]' + \theta_2\,[y_{1t-2}\ \ y_{2t-2}]' + [\eta_{1t}\ \ \eta_{2t}]' \tag{29.3}$$

Here, the θ_i are 2 by 2 matrices whose elements were estimated. We assumed the η_{it} were serially uncorrelated, and we produced estimates for the contemporaneous covariance matrix. Both OLS and FM-VAR techniques were used to produce parameter estimates.

As discussed above, FM-VAR estimation entails a kernel and bandwidth choice. We chose the Parzen kernel (Parzen 1957) because its use guarantees positive semi-definite covariance estimates (Andrews 1991) and because its use is common in the spectral estimation Monte Carlo literature (see, for example, Andrews 1991, Kitamura 1994, and Yamada and Toda 1996). We compared the results of three different bandwidth selection procedures. One was simply to fix the bandwidth at 3 for each sample magnitude. The other two methods were data-based. One method was taken from Andrews (1991) and the other from Schwert (1989).[27]

[27]More precisely, we used the AR(1) formula from Andrews 1991 (Equation 6.4, 835) to determine the Andrews automatic bandwidth. So chosen, the bandwidth can take any value on the positive real line and will generally vary from sample to sample. In our case, the values were smallest in

We compared the ability of different statistical tests to detect the existence (or absence) of Granger causal relationships between the variables in our system. In particular, we chose to compare the FM-VAR Wald test (henceforth, W-FM) with the more familiar OLS F test (henceforth, F-OLS) and the OLS Wald test (henceforth, W-OLS).[28] This choice was made for the following reasons. First, because FM theory is entirely asymptotic and because very little is known about the finite sample properties of FM-VAR estimators, it is necessary to appeal to asymptotic results when conducting hypothesis tests.[29] The Wald test was the natural choice as it was fully described and characterized for the FM environment by Phillips (1995). Given the decision to use W-FM, it is natural to also include W-OLS as a point of comparison. Finally, we included F-OLS because of its familiarity and because it has been used to generate many empirical results reported by political scientists and economists in published work.

We compared the size and power of F-OLS, W-OLS, and W-FM for Granger causality tests. To assess the *size* of each test, that is the tendency of the test to reject the null hypothesis when it is true, we proceeded as follows. First we set $b = 0$ in Equation 29.2. The parameters (a_1, a_2) took the values (1.0, 1.0), (0.95, 0.95), (0.90, 0.90), (0.50, 0.50), and (0.5, 1.0) in different experiments, corresponding to the full-rank integrated, nearly integrated, stationary, and mixed cases. We then estimated system (Equation 29.3) and tested the null hypothesis that y_{1t} does not Granger cause y_{2t}, that is, that $\theta_1(2,1) = \theta_2(2,1) = 0$. To determine the *power* of the test, or the probability that the test will reject a false null hypothesis, we generated data with the same set of values for a_1 and a_2, but set $b = 0.1$. We again estimated (Equation 29.3) and then tested the false null hypothesis that lags of y_{1t} do not help predict y_{2t}, or again that $\theta_1(2,1) = \theta_2(2,1) = 0$. All tests were conducted at the 5 percent nominal level. Empirical rejection probabilities were found by averaging over 1000 trials for each parameter configuration. Each data set was

the full-rank integrated, 25 observation case (typically between 1.0 and 2.0) and largest in the stationary, 500 observation environment (typically between 3.5 and 4.5).

The Schwert rule defines the bandwidth as an integer function of sample size. In our case, the bandwidths implied by the Schwert rule were 2, 3, 4, and 6 for sample sizes 25, 50, 100, and 500, respectively.

[28]If the estimated coefficients p have estimated covariance matrix S, then by a Wald test of the j linear restrictions $Rp = r$ we mean to determine the value $x^* = (Rp - r)'(RSR')^{-1}(Rp - r)$ and then compare x^* against the critical values from a $\chi^2(j)$ distribution. By an F-test we mean to compare the value x^*/j against the appropriate critical values from an $F(j,n)$ distribution, where n is the degrees of freedom from the sample used to generate the parameter estimates.

[29]The usual F statistic is a finite sample statistic which is appropriate in the case where one knows that the finite sample distribution of the estimator is approximately normal with a consistently estimated mean and variance. However, including lagged endogenous variables as regressors, as is required for Granger-causality regressions, renders the F test only asymptotically valid because the estimates themselves only have asymptotic properties (Hamilton, 1994, 304–5).

initialized with 100 presample observations, and these were not used in the experiments. This ensured the stationary components of the system, if any, had reached their long-run distributions before observations were taken. The coefficients were estimated with a corrected version of Gauss' COINT subroutine, FM-VAR.[30]

The results of the size tests are reported in Table 1, and the results of the power tests are reported in Table 2. The first column of each table describes the data generating process employed. The second column gives the sample magnitude. The third through seventh columns give the empirical rejection probabilities for each bandwidth selection rule. Note that the size results are almost invariant to the bandwidth, and the power results exhibit only modest variability in this regard. A few cases where these differences were significant will be noted below.

Consider the full-rank integrated process (row 1 of Table 1). In this case, testing the true null hypothesis that y_1 does not cause y_2 requires testing a nonstationary coefficient (namely, that on the first lag of y_1). It is well known that in such cases F-OLS and W-OLS will tend to over- reject, and our results again demonstrate that fact.[31] The tests were conducted at the 5 percent level, but the empirical sizes were around 11–12 percent for any number of observations for both OLS tests. On the other hand, note that the size of the W-FM test declined monotonically with the number of observations, from a high of 9 percent (twenty-five observations; Andrews 1991) to a low of 1 percent (500 observations, Fixed and Schwert). This finding is consistent with the theoretical result that in full-rank integrated environments, the size of W-FM tends to zero as the number of observations becomes large (see Phillips 1995, Theorem 6.3 and Remark 6.4, 1055). That is, the probability of a type I error, finding evidence of causality when none exists, is extremely small when the sample magnitude is large.

To clarify this idea we generated the empirical CDF of the W-FM statistic for the 100 observation, full-rank integrated case and plotted it against the CDF of a $\chi^2(2)$ random variable (see Figure 1). We plotted the $\chi^2(2)$ distribution because it is the 'bounding' distribution used to generate 'conservative' critical values for our W-FM hypothesis tests. From the figure, it is

[30]The original version of the FM-VAR subroutine was found to have a number of serious bugs. These were corrected with the help of one of the co-authors of COINT, Samuel Ouliaris. Copies of the new code are available from the authors. The Monte Carlo experiments were coded in MATLAB (version 4.2c.1). The MATLAB random number generator RANDN was used to generate the Gaussian disturbances to the data generating process; it was initialized with the seed 1029345 for each experiment. The programs were submitted to SUN Sparc10 workstations.

[31]For early work in this area see Granger and Newbold (1974). A more recent study of this problem is Ohanian (1988).

Table 1. Sizes Under Various Data Generating Process and Bandwidth Specifications*

Data Generating Process	Sample Size	F-OLS	W-OLS	W-FM Fixed	W-FM Andrews	W-FM Schwert
Full Rank Integrated	25	0.12	0.15	0.07	0.09	0.09
	50	0.11	0.13	0.03	0.04	0.03
Y1(t+l) = Y1(t) + e1(t+1)	100	0.11	0.13	0.03	0.04	0.03
Y2(t+l) = Y2(t) + e2(t+l)	500	0.11	0.11	0.01	0.02	0.01
Near Integrated	25	0.11	0.12	0.07	0.08	0.07
	50	0.08	0.09	0.03	0.03	0.03
Y1(t+1) = 0.95*Y1(t) + e1(t+1)	100	0.07	0.08	0.03	0.03	0.03
Y2(t+l) = 0.95*Y2(t) + e2(t+l)	500	0.05	0.05	0.01	0.01	0.01
Near Integrated	25	0.09	0.11	0.07	0.08	0.08
	50	0.07	0.08	0.03	0.04	0.03
Y1(t+1) = 0.90*Yl(t) + e1(t+1)	100	0.06	0.07	0.03	0.03	0.03
Y2(t+l) = 0.90*Y2(t) + e2(t+l)	500	0.05	0.05	0.01	0.02	0.02
'Mixed'						
H0: Y1 does not cause Y2						
	25	0.07	0.11	0.09	0.10	0.09
Y1(t+1) = 0.5*Y1(t) + e1(t+1)	50	0.06	0.07	0.05	0.05	0.05
Y2(t+1) = Y2(t) + e2(t+1)	100	0.05	0.06	0.03	0.04	0.04
	500	0.05	0.05	0.02	0.02	0.03
H0: Y2 does not cause Y1						
	25	0.08	0.10	0.05	0.06	0.06
Y1(t+1) = 0.5*Y1(t) + e1(t+1)	50	0.06	0.06	0.03	0.03	0.03
Y2(t+1) = Y2(t) + e2(t+1)	100	0.08	0.08	0.02	0.03	0.03
	500	0.06	0.06	0.02	0.02	0.02
Stationary	25	0.07	0.08	0.08	0.08	0.08
	50	0.05	0.06	0.05	0.05	0.05
Y1(t+1) = 0.50*Y1(t) + e1(t+1)	100	0.05	0.05	0.04	0.04	0.05
Y2(t+1) = 0.50*Y2(t) + e2(t+1)	500	0.04	0.04	0.03	0.03	0.04

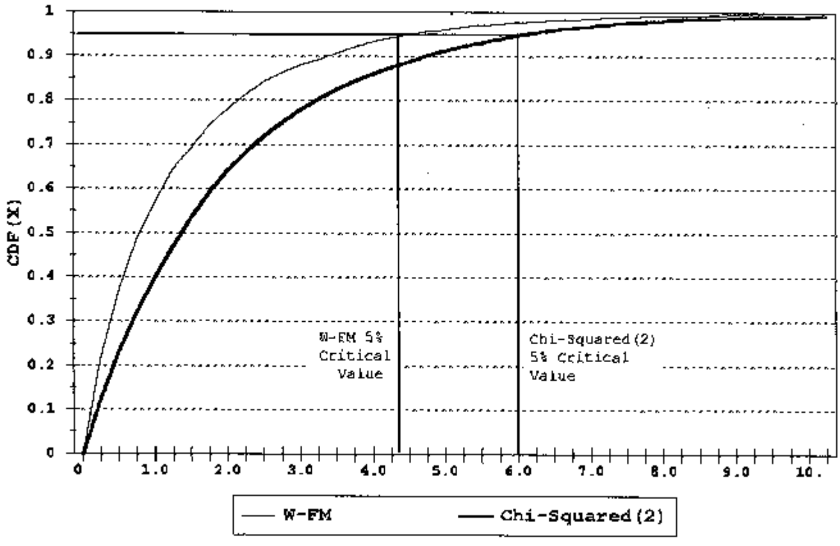
*Unless otherwise noted, the null hypothesis is H0: Y1 does not Granger cause Y2.

Table 2. Rejection Probabilities Under Various Data Generating Process and Bandwidth Specifications*

		Rejection Probabilities				
		F-OLS	W-OLS	W-FM		
Data Generating Process	Sample Size			Fixed	Andrews	Schwert
Full Rank Integrated	25	0.20	0.25	0.12	0.18	0.13
	50	0.44	0.52	0.12	0.23	0.12
Y1(t+1) = Y1(t) + e1(t+1)	100	0.87	0.88	0.33	0.49	0.28
Y2(t+1) = Y2(t) + 0.1Y1(t) + e2(t+1)	500	1.00	1.00	1.00	1.00	0.99
Near Integrated	25	0.21	0.26	0.14	0.16	0.15
	50	0.36	0.39	0.10	0.14	0.10
Y1(t+1) = 0.95*Y1(t) + e1(t+1)	100	0.63	0.65	0.13	0.22	0.12
Y2(t+1) = 0.95Y2(t) + 0.1Y1(t) + e2(t+1)	500	1.00	1.00	0.68	0.89	0.61
Near Integrated	25	0.19	0.24	0.12	0.14	0.13
	50	0.26	0.29	0.07	0.10	0.07
Y1(t+1) = 0.90*Y1(t) + e1(t+1)	100	0.46	0.47	0.09	0.17	0.09
Y2(t+1) = 0.90Y2(t) + 0.1Y1(t) + e2(t+1)	500	0.99	0.99	0.48	0.69	0.49
'Mixed'	25	0.10	0.14	0.11	0.12	0.10
	50	0.11	0.12	0.08	0.08	0.08
Y1(t+1~ = 0.5*Y1(t) + e1(t+1)	100	0.16	0.17	0.09	0.10	0.10
Y2(t+1) = Y2(t) + 0.1*Y1(t)+e2(t+1)	500	0.61	0.61	0.32	0.37	0.41
Stationary	25	0.11	0.14	0.11	0.11	0.11
	50	0.10	0.12	0.08	0.07	0.08
Y1(t+1) = 0.50*Y1(t) + e1(t+1)	100	0.15	0.16	0.10	0.10	0.12
Y2(t+1) = 0.50Y2(t) + 0.1Y1(t) + e2(t+1)	500	0.62	0.62	0.40	0.48	0.54

*The null hypothesis is H0: Y1 does not Granger cause Y2.

Figure 1. Empirical CDF of FM-VAR Wald Statistic for I(1) Specification



clear that the true 5 percent critical value for the W-FM statistic is around 4.3 (for this specific case). The reason that W-FM displays sizes smaller than 5 percent, then, is that the bounding distribution's critical value of 5.99 actually is used in the hypothesis tests, instead of the smaller true critical value.

The nearly-integrated cases are reported in rows two and three of Table 1. Although the environment is stationary, W-FM still rejects the Granger causality null hypothesis less than 5 percent of the time. For instance, with 500 observations the empirical size never exceeds 2 percent. Observe also that F-OLS and W-OLS still tend to over-reject the (true) null hypothesis at all but the largest number of observations. In the case $a_i = 0.95$, samples as large as 100 still produce F-OLS and W-OLS tests that reject the Granger causality hypothesis 7 percent of the time at the nominal (5 percent) level. This size distortion arises because of the very complicated small-sample distributions of the estimators in near integrated environments.

Next, consider the mixed system (row four of Table 1). This system is not symmetric; the coefficients on the y_1 process are stationary, while those on the first lag of y_2 are nonstationary. We, therefore, tested two hypotheses: that y_1 does not cause y_2, and that y_2 does not cause y_1. As expected, the OLS tests performed very well when stationary coefficients are restricted and slightly less well otherwise. For example, with sample magnitude 100 the empirical size of F-OLS was identical to its nominal size, 5 percent when the stationary coefficients were tested, but 8 percent when the nonstationary coefficients were tested. Similarly, the sizes of W-FM are larger in the stationary case, especially when there are few observations in the sample.

The last row of Table 1 summarizes the results for the stationary case. As expected, F-OLS and W-OLS have sizes that very closely resemble their levels. F-OLS, of course, performs slightly better when the sample magnitude is not large. W-FM exhibits sizes that are very close to those of the OLS tests. This result was expected. It is consistent with the theoretical finding that in stationary environments W-OLS and W-FM are equivalent.

Finally, consider briefly Table 2, which details the results of our power experiments. Given the size results, the outcome of the power tests are not surprising. In particular, we saw that the empirical size of W-FM was much smaller than F-OLS or W-OLS in integrated and near-integrated environments. Thus, as expected, the power of W-FM is also smaller than F-OLS or W-OLS in those environments. For instance, for 100 observations in the integrated environment F-OLS accurately finds evidence of causality in 87 percent of the time, while W-FM finds such evidence in as few as 28 percent of the cases (Schwert bandwidth) and no more than 49 percent of the cases (Andrews bandwidth). It is interesting to note, however, that W-FM also has less power than F-OLS and W-OLS when the environment is stationary. From the last row of Table 2, we see that, for all sample magnitudes and bandwidth selection rules, W-FM rejects the null a smaller number of times than the OLS tests, and the difference does not seem to diminish as the sample magnitude increases. When the sample magnitude is 500, the OLS tests reject the null of noncausality in 62 percent of the trials, while W-FM does so 40 percent of the time (fixed bandwidth) and 54 percent of the time (Schwert bandwidth).

Overall, the results of our Monte Carlo experiment are as expected. The FM-VAR Wald test is attractive in that it has consistently smaller sizes than either the OLS F-test or the OLS Wald test. On the other hand, the OLS tests had greater power. Both results were robust to the bandwidth selection rule employed. Together, the results revealed the 'conservative' nature of the FM-VAR approach to hypothesis testing. In particular, for Granger causality tests, the FM-VAR Wald test statistic was biased toward accepting the null hypothesis (no causality) when its value was compared against a critical value from the appropriate bounding χ^2 distribution. The stage is, therefore, set for a reevaluation of some existing, level VARs in political science.

4.2 The Arms Race Debate Revisited: India's Culpability Reconsidered

The late 1970s witnessed much debate about the existence and nature of so-called "arms races." International relations scholars were convinced that many countries were locked in dangerous, mutually reinforcing arms buildups. But they were unable to show this joint causality in the relevant time series data (for instance, see Majeski and Jones 1980). Freeman (1983) argued that the confusion derived from the use of inappropriate statistical methods. Using a direct Granger test in the context of what was essentially a VAR model in differences, Freeman found that there were robust Granger causal relationships between Indian and Pakistani arms spending series, namely, Indian arms spending Granger caused Pakistani arms spending but not vice versa. Interestingly, current accounts of Indian-Pakistani relations also express concern about a nuclear arms buildup provoked by political developments within India (*The Economist*, May 4, 1996: 35–7).

Freeman's inferences could be flawed in several ways. First, he used differences in order to reduce serial correlation without any concern for the effect this would have on the long-run relationship between arms spending processes. In addition, in order to show the robustness of his results, Freeman added linear deterministic trends to his equations without knowing how this might change the limiting distribution of his estimator if the arms expenditure series had unit roots. This possibility of unit roots (cointegration) was never considered by Freeman. In fact, since he wrote, both theoretical (Williams and McGinnis 1988) and experimental studies (Rajmaira 1995) have appeared which suggest the existence of unit roots and(or) cointegration in India and Pakistan spending data.[32] For all these reasons, Freeman's study is prone to the problems discussed in part one of this paper.

The Indian-Pakistani arms buildup was reanalyzed in the following way. First, a comparable data set was constructed from the original source for the same period, 1948–1975.[33] Then a standard, level VAR model of the Indian

[32] Freeman (1983) did not explicitly build a VAR model. He ran direct Granger tests on systems of equations with two lagged endogenous terms and two and four lagged (presumably) exogenous terms. He logged and differenced the Majeski and Jones data. He also used a Zellner-Aitken correction for correlation of the errors in the equations. What we produce here then is essentially a level VAR benchmark: what Freeman would have found had he built the conventional model. As noted in the text, the results are qualitative identical to those which Freeman reports in Table 4 (1983, 352–3) of his article.

For events time series in the period 1980–1990, Rajmaira (1995, especially 16) finds much evidence of unit roots and cointegration in Indian and Pakistani foreign policy behavior.

[33] It has been 15 years since the Majeski and Jones study appeared. Freeman no longer has the data set those authors sent to him. And Majeski (personal communication) also does not have it. So we tried to construct the series from the original source cited by Majeski and Jones. We found that

and Pakistani arms spending system was constructed (without any provision for the possibility of nonstationarity). Two lags of the variables, a constant, and no time trends were included in the respective equations. As in the Freeman study, natural logs of the series were used. This yielded qualitative results identical to Freeman's: the F-OLS test indicated that Indian arms spending Granger caused Pakistani arms spending but not vice versa; the test of the hypothesis that Indian arms expenditures do not Granger cause Pakistani arms expenditures had a F-OLS p value of .01, and the hypothesis that Pakistani arms expenditures do not Granger cause Indian arms expenditures had a F-OLS p value of .65 (Table 3).

Next, the system was estimated with FM-VAR. Because FM techniques have not yet been used in political science it is useful at this point to describe the way in which the estimates are obtained in this particular case. This specificity will help make more concrete the necessarily abstract discussion pursued above. It will also help highlight the importance of bandwidth and kernel selection to the estimation procedure.

The practitioner's practical goal is to obtain values for the elements associated with Equation 28, which is repeated here for convenience (slightly modified to take account of the model's constant, which is a column of ones denoted by ι).

$$\hat{F}^{+} = [Y'Z \,\vdots\, Y'Y_{-1} - \hat{\Omega}_{\varepsilon Y}\hat{\Omega}_{yy}^{-1}(\Delta Y'_{-1}Y_{-1} - T\hat{\Delta}_{\Delta Y \Delta Y}) \,\vdots\, Y'\iota](X'X)^{-1} \quad (30)$$

Within the context of the Indian-Pakistani analysis, we proceed as follows. The matrix $\mathbf{Y}$ is simply the two-column data matrix. The first column is the data for India, and the second is the data for Pakistan, so that $\mathbf{Y} = [\mathbf{y}_{IND}, \mathbf{y}_{PAK}]$. Because in this analysis we include only two lags, the matrix $\mathbf{Z}$ is given by $\mathbf{Z} = [(\Delta\mathbf{y}_{IND})_{-1}, (\Delta\mathbf{y}_{PAK})_{-1}]$. The matrix $\mathbf{X}$ is then given by $\mathbf{X} = [\mathbf{Z}, \mathbf{Y}_{-1}, \iota]$, where, again, ι is a column of ones. Notice that the VAR in levels with two lags and a constant can simply be expressed as $\mathbf{Y} = \mathbf{XF}' + \mathbf{E}$, where $\mathbf{F}'$ is a 5 by 2 coefficient matrix. It is useful to let $\hat{\mathbf{E}} = \mathbf{Y} - \mathbf{X}\hat{\mathbf{F}}'$ be the typical OLS residuals associated with the level VAR regression.

It remains to define the one-sided and two-sided long-run covariance matrices. Although the notation is confusing (we have adopted the Phillips (1995) notation), the estimated matrix $\hat{\Omega}_{\varepsilon Y}$ is the two sided long-run co-

those authors apparently never corrected for the use of different price deflators by the Stockholm Peace Research Institute between 1948 and 1975. We used the series for the 1973 price deflator for the observations for the years 1954–1975. For the remaining five observations (1948–1953) we converted the 1960 price data to data in 1973 prices. Our lag-length tests (available on request) indicate that two lags of each rhs variable are called for.) Again, for our naive level VAR, the results were qualitatively identical to those reported in Freeman (1983).

Table 3. India-Pakistan Causality Tests

Equation	Block of Coefficients	p-value*	
		F-OLS	W-FM
India	India	0.00	0.00
	Pakistan	0.65	0.53
Pakistan	India	0.01	0.00
	Pakistan	0.00	0.00

*W-FM p-values are upper bounds for the 'true' values.

variance between $\hat{E}$ and $\Delta\mathbf{Y}_{-1}$. Similarly, the estimated matrices $\hat{\Omega}_{YY}$ and $\hat{\Delta}_{\Delta Y \Delta Y}$ are, respectively, the two-sided and one-sided long run covariance matrices for $\Delta\mathbf{Y}_{-1}$. The covariance matrices are estimated using well known nonparametric techniques, as we next briefly describe.

In general, if u_t and v_t are (not necessarily distinct) vector processes we can form kernel estimates of their two and one sided long-run covariance matrices (see, e.g., Priestley 1981). These have form

$$\hat{\Omega}_{uv} = \Sigma_{j=-T+1,T-1}\omega(j/K)\hat{\Gamma}(j) \tag{31}$$

$$\hat{\Delta}_{uv} = \Sigma_{j=0,T+1}\omega(j/K)\hat{\Gamma}(j) \tag{32}$$

The values $\hat{\Gamma}$ are just the sample covariances and are given by

$$\hat{\Gamma}(j) = (1/T)\Sigma_{1\le t,t+j\le T}u'_{t+j}, v_t \tag{33}$$

The weights $\omega(\bullet)$ are given by a kernel function, and the value $K \ge 0$ is the bandwidth parameter. The bandwidth value defines the number of lagged covariance terms to be included in the estimation of the long-run matrices. In particular, kernels are usually defined so that $w(j/K) = 0$ whenever $|j/K| \ge 1$.

We chose to use the Parzen kernel for our empirical work. As mentioned earlier, the Parzen kernel has been frequently used in covariance matrix estimation and has attractive properties (see, e.g., Phillips 1995, 1031). The Parzen kernel is defined by

$$\omega(x) = \begin{cases} 1 - 6x^2 - 6|x|^3 & \text{if } 0 \le x \le 1/2 \\ 2(1 - |x|)^3 & \text{if } 1/2 \le x \le 1 \\ 0 & \text{otherwise} \end{cases} \tag{34}$$

To determine the value for the bandwidth parameter we used the Andrews (1991) data based automatic bandwidth estimator for the Parzen kernel (his equation 6.2, 834). The value of four was selected by this estimator.

The results of the estimation are found in the far right hand column of Table 3. The same lag-length specification was used in this system; the bandwidth parameter was set by an automatic routine at 4. The results are reported in the far right-hand columns of the table. They indicate that the p value associated with the null hypothesis that Indian arms spending does not Granger cause Pakistani spending is as great as .00, while the p value associated with the null that Pakistani arms spending does not Granger cause Indian arms spending is as great as .53. Hence the implication is that India is indeed responsible for the arms race in the indicated period. The Indian-Pakistani dyad still is characterized by one-way causation when the possibility of unit roots and cointegration is taken into account. In view of the results of our Monte Carlo experiment, we are relatively confident that this result is accurate, despite the relatively short lengths of our time series in this case.

4.3 Approval and Expectations: A Reevaluation

Virtually every student of political economy believes that there is a direct connection between the state of the economy and presidential popularity. Recently, however, MacKuen, Erikson, and Stimson (1992) questioned this, arguing that the linkage between the economy and approval is indirect and mediated by aggregate economic perceptions about the economy (in particular, about expectations of future business conditions). They claim that economic conditions are exogenous to economic perceptions and that economic perceptions are then exogenous to presidential approval. The evidence for this is a set of Granger causality tests on the respective system of time series; these tests define what they call the basic "causal web" governing the economy and presidential approval. On the basis of these Granger causality results, MacKuen, Erikson, and Stimson proceed to construct a recursive structural model for approval, a model which assumes no simultaneity between business expectations and approval.

Our central concern is the ways in which unit roots in time series data undermine Granger causality tests. This concern is obviously relevant for MacKuen, Erikson, and Stimson (1992). The authors do not test for the presence of unit roots in any of their time series. However, others who have studied presidential approval and inflation (Ostrom and Smith 1993, Table 2) and economic expectations (Durr 1993b) have concluded that the relevant time series do, in fact, contain unit roots. For these reasons, MacKuen, Erikson, and Stimson's results may suffer from simultaneity bias.

But there are other reasons to question the veracity of MacKuen, Erikson, and Stimson's results. There are both theoretical and empirical reasons to suspect that approval drives economic expectations. For example, in his analysis of the relationship between economic expectations and policy mood, Durr (1993b) does not simply use raw economic expectations to predict policy mood. Rather, he first regresses economic expectations on four indicators of the objective economy. (He does this to isolate the variance in that series that truly represents economic expectations, as opposed to other forces.) But the residuals from that regression are not random.

> Judging from the pattern exhibited [in the residuals], it is certain that something far more systematic than random errors of sampling or perception is at work. Indeed, the residuals suggest the existence of a *political* component of long-term business expectations. Peaks in the time series correspond well to presidential elections, suggesting that consumer sentiment, like presidential approval, is vulnerable to a honeymoon effect. (Durr 1993b, 161, italics in original)

This provocative finding suggests the possibility that economic expectations are, to some extent, a *function* of presidential approval, rather than (or in addition to) a *cause* of approval, as MacKuen, Erikson, and Stimson argue. This possibility, while certainly less flattering a portrait than the "bankers" that MacKuen, Erikson, and Stimson describe, is very real.

With the help of MacKuen, Erikson, and Stimson, we reconstructed their dataset.[34] With this data and using the authors' specification, we replicated the qualitative findings of their Table 1.[35] Specifically, although the exact p values from our estimation are not identical to theirs, the results of all of our causality tests are identical to those which MacKuen, Erikson, and Stimson report.

We then estimated a five-variable system composed of presidential approval, business expectations, personal expectations, the unemployment rate, and the consumer price index as the endogenous variables. We chose these five variables because they are central to MacKuen, Erikson, and Stimson's causal claims that economic conditions (unemployment and the cpi) cause economic perceptions (personal and business expectations), which, in turn, cause approval.[36]

[34] Unfortunately, the authors' original dataset no longer exists. MacKuen, Erikson, and Stimson were extremely helpful in recreating the data. Their assistance is gratefully acknowledged.

[35] The replication table is not presented here. It is available from the authors upon request.

[36] All of the results reported here include six deterministic variables: a dummy variable for a series of events (detailed in MacKuen, Erikson, and Stimson's endnote 5); a variable for the buildup to the Vietnam war (detailed in MacKuen, Erikson, and Stimson's endnote 3); and four dummy variables for the first four quarters of a new administration. However, both the need for four lags as well as all causal findings remain unchanged if these variables are omitted.

Before proceeding to the results, the issue of lag length specification must be acknowledged. MacKuen, Erikson, and Stimson never test for the appropriate lag length specification. Rather they simply posit one lag for their endogenous variable and two lags for the other variables on the right-hand side of their equations. When we tested for the appropriate lag specification for our five-variable system, we found that four lags of all right-hand side variables are called for. Therefore, it is clear from the outset that MacKuen, Erikson, and Stimson's Granger causality analysis suffers from omitted variables bias (see Freeman 1983). Because we are primarily interested here in the problems caused by unit roots, we worked with the four-lag system for our comparison of the relevant F-OLS and W-FM statistics.

These results are presented in Table 4. The OLS results show that, consistent with MacKuen, Erikson, and Stimson's claims, the objective economy has no direct effect on approval; the respective p values for the unemployment rate and the cpi are .96 and .95. Also, expectations of future business conditions drive approval (p = .02), but the reverse is not the case (p = .29). The OLS results, then, are exactly consistent with the MacKuen, Erikson, and Stimson findings. Of course, these results do not take into account any trends or cointegration that might be present in this system. For these, we need to examine the second column of Table 4. The W-FM statistics indicate that the p value associated with the null that business expectations do not Granger cause approval is as great as .10, and the p value associated with the null that approval does not Granger cause business expectations is as great as .03[37]. In other words, the W-FM results indicate that the MacKuen, Erikson, and Stimson recursive system is incorrectly specified. Once the possibility of unit roots and cointegration is entertained, their causal findings are reversed using this method. The FM-VAR results suggest that economic expectations do not drive approval; rather, approval drives expectations.[38] In light of this,

[37]On a substantive note, however, it is important to mention that the sum of the approval coefficients in the economic expectations equation are negative, which is exactly the opposite of what would be expected. That is, high levels of approval lead to *decreases* in economic expectations, and low approval leads to net *increases* in expectations. Note that this result is consistent with those that Williams (1990) obtained. Williams found that high approval allows policymakers to adopt stringent monetary policies and to make other controversial choices they could not make if approval were low. The critical point is that MacKuen, Erikson, and Stimson's results are in error.

[38]A technical note: In our equations, we did not difference the unemployment rate or the consumer price index, as most scholars have done. Such arbitrary detrending naturally defeats the purpose of FM-VAR, the goal of which is to specify systems without pretesting or detrending. Differencing, of course, is an implicit statement about the trend properties of a variable—specifically, that the variable contains a unit root. That said, we should note that the cpi, when left undifferenced, grows exponentially. Including this variable in the system made bandwidth selection a sensitive issue. Therefore, in order to test the robustness of our results, we tested a four-variable system that excluded the cpi (not shown). The results differed somewhat, with the key finding being that approval and economic expectations were mutually endogenous. While this leads to somewhat different substantive conclusions than those presented in Table 4, such findings still undermine the

Table 4. Approval and Expectations: A Re-evaluation

Equation	Block of Coefficients	p-value* F-OLS	p-value* W-FM
APPROVAL	APPROVAL	0.00	0.00
	UNEMPLOYMENT RATE	0.96	0.86
	CPT	0.95	0.75
	PERSONAL EXPECTATIONS	0.84	0.80
	ECONOMIC EXPECTATIONS	0.02	0.10
UNEMPLOYMENT RATE	APPROVAL	0.69	0.44
	UNEMPLOYMENT RATE	0.00	0.00
	CPT	0.36	0.04
	PERSONAL EXPECTATIONS	0.79	0.57
	ECONOMIC EXPECTATIONS	0.46	0.05
CPT	APPROVAL	0.87	0.51
	UNEMPLOYMENT RATE	0.00	0.00
	CPT	0.00	0.00
	PERSONAL EXPECTATIONS	0.21	0.35
	ECONOMIC EXPECTATIONS	0.79	0.86
PERSONAL EXPECTATIONS	APPROVAL	0.69	0.40
	UNEMPLOYMENT RATE	0.30	0.03
	CPT	0.11	0.33
	PERSONAL EXPECTATIONS	0.00	0.00
	ECONOMIC EXPECTATIONS	0.08	0.13
ECONOMIC EXPECTATIONS	APPROVAL	0.29	0.03
	UNEMPLOYMENT RATE	0.06	0.01
	CPT	0.28	0.15
	PERSONAL EXPECTATIONS	0.64	0.08
	ECONOMIC EXPECTATIONS	0.00	0.00

*W-FM p-values are upper bounds for the 'true' values.

the MacKuen, Erikson, and Stimson models that specify economic expectations to be exogenous to approval, therefore, may be misspecified, and their conclusions should be reconsidered.

4.4 Summary

These reanalyses suggest that to some extent, level VAR results are unsound. Some of the respective sets of causal inferences are plagued by the failure to take into account the possibility of unit roots and cointegration. Inferences about the causes of certain arms buildups remain the same when

FM-VAR is applied. However, the use of FM-VAR suggests that MacKuen, Erikson, and Stimson may have misspecified the causal relationships of the determinants of presidential approval. Because they did not take into account the possibility of unit roots and cointegration in their time series, those authors may have obtained inaccurate results from their causality tests and, in turn, posited a recursive structure for their model when such a structure, in fact, was not supported by their data. Clearly, the dynamics of approval and perceptions about the economy deserve further study along these lines.

5. Conclusion

The primary conclusion here is that political scientists must pay attention to the trend properties of their data. Conceptually, we must be clearer about why particular political processes are characterized by one type of trend rather than another. Theoretical progress must be made in deriving the likely trend properties of various political processes. Illustrative is the article by Williams and McGinnis (1988) that showed how a well-established theory of arms races implied random-walk behavior on the part of the superpowers. More work of this kind must be done by political theorists. For example, formal modelers need to be clearer about exactly what their models imply insofar as trend properties are concerned and then subject those implications to empirical tests.[39] Statistically, the impact of unit roots and cointegration on the innovation accounting in published works needs to be explored (cf. Phillips forthcoming). And the usefulness of various model-selection methods in this context ought to be examined. In addition, further study of FM-VAR and its performance relative to other approaches certainly is called for. As regards the latter task, Sims, Stock, and Watson's approach and its successors (Toda and Phillips 1993) is exceedingly complicated; no general software exists for it. The Johansen method is more feasible. Researchers simply have to keep in mind the possibility of pretest bias and other problems outlined above. There are still other, newer approaches appearing, such as the overfit method (Toda and Yamamoto 1995; Dolado and Lutkepohl 1996). These methods ought to be studied as well. Among these challenges for political science, we feel that the first two are more pressing than the third; statistical work is ahead of conceptual and theoretical work in political science, in our opinion.

As regards time-series modeling, at this point, political scientists have several options. FM-VAR is an appealing method for several reasons. First,

MacKuen, Erikson, and Stimson assertion that expectations drive approval (but not the reverse). Surely this study will not be the last word on the relationship between these two variables; rather, we hope that it leads to subsequent investigations that carefully consider what variables belong in the system, and allow for the possibility of trends and cointegration in the data.

[39] A very useful (if analytically challenging) contribution in economics here is Singleton (1988).

it does not require exact knowledge of the trend properties of one's time series data. Second, Monte Carlo analysis shows that FM-VAR outperforms its OLS counterpart in important respects, principally size. Hence it can be used to perform stringent causality tests, where the risk of a Type I error will be minimal. On the other hand, as we have noted, FM-VAR has some shortcomings. For example, FM-VAR produces the upper bounds for p values for tests of non-Granger causality. (And if one believes that statistical inference already errs too much of the side of Type I error, exact p values may be desired, if difficult to obtain.) In particular, the risks of Type II errors are potentially larger with FM-VAR than with its OLS counterpart; findings that exist in reality may go undetected. Individual scholars may wish to choose the method that best fits the particular purposes of their research. Also, there remain several important research questions regarding the method, including how best to estimate the long-run covariances that are at the heart of this estimator (Cappuccio and Lubian 1995; see also Yamada and Toda 1996).

Some political scientists may want to continue pretesting data, using unit root tests to decide whether or not to difference data or to employ an error-correction transformation, and ultimately applying the Sims, Stock, and Watson, or Johansen approaches to causality testing. It is important to reiterate several things about this strategy. First, the respective tests are always subject to mistaken inference about the existence of unit roots and cointegration. Moreover, because the null of many of these tests is that a unit root is present, the "pretest mindset" may lead to inappropriate differencing of time series. Time series can be more or less cointegrated, just like series can have a more or less long-term component. Using levels does not throw out this nearly-integrated component like differencing does. Third, if one prefers the functional form which students of unit roots recommend, the VECM, the analyst can simply posit it as a reduced form whether or not there is evidence of integrated regressors (Beck 1993). In fact, FM-VAR estimation employs a functional form which is easily transformed into a VECM formulation (see note 20). Last, the causality tests to which pretesting leads, especially the Sims, Stock, and Watson method, is, as we have argued, every bit as complex (if not more complex) than FM-VAR. The fact that no political scientist who advocates pretesting apparently has ever used the Sims, Stock, and Watson or Johansen approaches to check the accuracy of existing causal inferences testifies to this.

A third approach is to analyze our data in levels. If we do this and our series do not contain cointegration and (or) unit roots, our causality tests and innovation accounting are sound. If there are cointegrated series and (or) unit roots in the system, the moving average responses and the decomposition of error variance will still have some descriptive value. But, our statistical tests will be biased against the null of Granger non-causality. In fact, roughly 20 percent of our tests will inaccurately reject this null (Ohanian

1988). Some econometricians suggest correcting these values through simulation (Christiano and Lungqvist 1988). Others suggest that the original test statistics be used, noting that under specific Bayesian assumptions, p values (or probability values in this case) are fine (Sims 1988; Sims and Uhlig 1991). And even though Bayesian theory behind such an analysis is questionable (Phillips 1991a,b), the behavior of the likelihood is undeniably described by test statistic values. Thus, if we break out of the frequentist mold—something only a few political scientists have done (Williams 1993)—we can analyze our data in levels. Also note that error-correction VARs have an advantage if all variables are cointegrated. If not, this approach is asymptotically equivalent to a VAR in levels. The choice of a VECM is in this sense a matter of taste.[40]

In sum, we believe that all these strategies have some scientific value. If each recognizes the costs and benefits associated with it, each strategy can make real contributions. FM-VAR users must address the inability to compute exact p values. Those who favor pretesting must admit that Type II error is a major problem and not fall back on significance tests. Such scholars must also apply the more challenging Sims, Stock, and Watson or Johansen tests eventually to demonstrate the implications of unit roots and cointegration for causality testing, something that has not yet been done with political data. VAR analysts who estimate in levels need to be keenly aware that their test statistics may not have the typical interpretation. They should not interpret the results of the tests on nonstationary coefficients without persuasive evidence that the respective inferences are legitimate. In the end, all these caveats and conundrums are beneficial because they require thought from time series analysts. The contribution of FM-VAR is that it moves this thought forward, making it possible for the first time to assess the soundness of our causal inferences in the face of unit roots and cointegration.

Manuscript submitted 22 May 1997.
Final manuscript received 22 December 1997.

REFERENCES

Alesina, A., and H. Rosenthal. 1995. *Partisan Politics, Divided Government and the Economy.* New York: Cambridge University Press.

Andrews, Donald W. K. 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation." *Econometrica* 59:817–58.

[40] Our focus in this paper is on Granger causality tests, which are widely used in political science. Some VAR analysts rely less heavily on these tests and more heavily on the analysis of the moving average representation with Bayesian posterior distributions; these distributions are easy to compute (see Sims 1987, 1988). It should be possible for FM-VAR to be used to generate confidence bands for those uncomfortable with this Bayesian approach.

Beck, Nathaniel. 1993. "The Methodology of Cointegration." *Political Analysis* 4:237–48.

Box-Steffensmeier, Janet, and Renée M. Smith. 1996. "Investigating Political Dynamics Using Fractional Integration Methods." Unpublished manuscript. Ohio State University.

Cappuccio, Nunzio, and Diego Lubian. 1995. "Fully Modified Estimation of Cointegrating Vectors via VAR Prewhitening: A Simulation Study." Unpublished manuscript. University of Padova.

Christiano, Lawrence, and Lars Ljungqvist. 1988. "Money Does Granger-Cause Output in a Bivariate Money-Output Relation." *Journal of Monetary Economics* 22:217–35.

Chrystal, K. Alec, and David A. Peel. 1986. "What Does Economics Have to Teach Political Science and Vice Versa?" *American Economic Review* 76:62–5.

DeJong, David N., John C. Nankervis, N. E. Savin, and Charles H. Whiteman. 1992. "Integration Versus Trend Stationarity in Time Series." *Econometrica* 60:423–33.

Dixit, Avinash K., and Robert S. Pindyck. 1994. *Investment Under Uncertainty*. Princeton: Princeton University Press.

Dolado, J. J., and Helmut Lutkepohl. 1996. "Making Wald Tests for Cointegrated VAR Systems." *Econometric Reviews* 15:369–86.

Durr, Robert H. 1993a. "An Essay on Cointegration and Error Correction Models." *Political Analysis* 4:185–228.

Durr, Robert H. 1993b. "What Moves Policy Sentiment?" *American Political Science Review* 87:158–70.

Engle, Robert F., and S. Yoo. 1991. "Cointegrated Economic Time Series: An Overview with New Results." In *Long-Run Economic Relationships: Readings in Cointegration*, eds. Robert F. Engle and Clive W. J. Granger. Oxford: Oxford University Press.

Evans, G. B. A., and N. E. Savin. 1981. "Testing for Unit Roots: 1." *Econometrica* 49:753–79.

Evans, G. B. A., and N. E. Savin. 1984. "Testing for Unit Roots: 2." *Econometrica* 52:1241–69.

Freeman, John R. 1983. "Granger Causality and the Time Series Analysis of Political Relationships." *American Journal of Political Science* 27:327–58.

Freeman, John R., and James E. Alt. 1994. "The Politics of Public and Private Investment in Britain." In *The Comparative Political Economy of the Welfare State*, eds. T. Janoski and A. M. Hicks. New York: Cambridge University Press.

Freeman, John R., and Daniel Houser. 1998. "A Computable Equilibrium Model for the Study of Political Economy." *American Journal of Political Science*. 42:628–660.

Freeman, John R., John T. Williams, and Tse-Min Lin. 1989. "Vector Autoregression and the Study of Politics." *American Journal of Political Science* 33:842–77.

Fuller, Wayne A. 1976. *Introduction to Statistical Time Series*. New York: Wiley.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street: Strategic Reciprocity in World Politics*. Chicago: University of Chicago Press.

Goldstein, Joshua S., and John R. Freeman. 1991. "U.S.-Soviet-Chinese Relations: Routine, Reciprocity, or Rational Expectations?" *American Political Science Review* 85:17–35.

Granato, James, and Renée Smith. 1994. "Exogeneity, Inference, and Granger Causality: Part II: The Case of Integrated Regressors." *The Political Methodologist* 6:23–5.

Granger, C. W. J., and P. Newbold. 1974. "Spurious Regressions in Econometrics." *Journal of Econometrics* 2:111–20.

Hall, Robert. 1978. "Stochastic Implications of the Life Cycle-Permanent Income Hypothesis." *Journal of Political Economy* 86:971–87.

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

Hargreaves, Colin. 1994. "A Review of Methods of Estimating Cointegrating Relationships." In *Nonstationary Time Series Analysis and Cointegration*, ed. Colin P. Hargreaves. Oxford: Oxford University Press.

Harvey, Andrew C. 1990. *The Econometric Analysis of Time Series*. Cambridge, Mass.: MIT Press.

Hendry, D. F., and J. F. Richard. 1983. "The Econometric Analysis of Economic Time Series." *International Statistical Review* 51:111–63.

Johansen, S. 1988. "Statistical Analysis of Cointegrating Vectors." *Journal of Economic Dynamics and Control* 12:231–54.

Juselius, Katarina, and Colin W. Hargreaves. 1992. "Long-Run Relations in Australian Money Demand." In *Macroeconomic Modeling of the Long Run*, ed. Colin W. Hargreaves. London: Edward Elgar.

Kitagawa, Genshiro. 1987. "Non-Gaussian State-Space Modeling of Nonstationary Time Series." *Journal of the American Statistical Association* 82:1032–63.

Kitamura, Yuichi. 1994. "Hypothesis Testing in Models with Possibly Nonstationary Processes." Unpublished manuscript. University of Minnesota.

Le, Nhu D., Adrian E. Raftery, and R. Douglas Martin. 1996. "Robust Bayesian Model Selection for Autoregressive Processes With Additive Outliers." *Journal of the American Statistical Association* 91:123–31.

Lee, Chin Chin, and Peter C. B. Phillips. 1994. "An ARMA Prewhitened Long-Run Variance Estimator." Unpublished manuscript.

MacKuen, Michael B., Robert S. Erikson, and James A. Stimson. 1992. "Peasants or Bankers? The American Electorate and the U.S. Economy." *American Political Science Review* 86:597–611.

Majeski, Stephen J., and David Jones. 1980. "Arms Race Modeling: Causality and Model Specification." *Journal of Conflict Resolution* 25:259–88.

Malliaris, A. G., and W. A. Brock. 1982. *Stochastic Methods in Economics and Finance*. New York: North-Holland.

Martin, R. Douglas. 1980. "Robust Estimation of Autoregressive Models." Unpublished manuscript. University of Washington.

Nelson, Charles R., and Heejoon Kang. 1981. "Spurious Periodicity in Inappropriately Detrended Time Series." *Econometrica* 49:741–51.

Nelson, Charles R., and Charles I. Plosser. 1982. "Trends and Random Walks in Macroeconomic Time Series." *Journal of Monetary Economics* 10:139–62.

Nickell, S. 1985. "Error Correction, Partial Adjustment and All That: An Expository Note." *Oxford Bulletin of Economics and Statistics* 47:119–29.

Ng, Serena, and Pierre Perron. 1995. "Unit Root Tests in ARMA Models With Data-Dependent Methods for the Selection of the Truncation Lag." *Journal of the American Statistical Association* 90:268–80.

Ohanian, Lee E. 1988. "The Spurious Effects of Unit Roots on Vector Autoregressions: A Monte Carlo Study." *Journal of Econometrics* 39:251–66.

Ostrom, Charles W., and Renée M. Smith. 1993. "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." *Political Analysis* 4:127–83.

Parzen, E. 1957. "On Consistent Estimates of the Spectrum of a Stationary Time Series." *Annals of Mathematical Statistics* 28:329–48.

Phillips, Peter C. B. 1988a. "Weak Convergence of Sample Covariance Matrices to Stochastic Integrals via Martingale Approximations." *Econometric Theory* 4:528–33.

Phillips, Peter C. B. 1988b. "Weak Convergence to the Matrix Stochastic Integral $\int_0^1 BdB'$." *Journal of Multivariate Analysis* 24:252–64.

Phillips, Peter C. B. 1991a. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends." *Journal of Applied Econometrics* 6:333–64.

Phillips, Peter C. B. 1991b. "Optimal Inference in Cointegrated Systems." *Econometrica* 59:238–306.

Phillips, Peter C. B. 1992a. "Hyperconsistent Estimation of a Unit Root in Time Series Regression." Cowles Foundation Discussion Paper No. 1040.

Phillips, Peter C. B. 1992b. "Simultaneous Equation Bias in Level VAR Estimation." *Econometric Theory* 8:307.

Phillips, Peter C. B. 1995. "Fully Modified Least Squares and Vector Autoregression." *Econometrica* 63:1023–78.

Phillips, Peter C. B. n.d. "Impulse Response and Forecast Error Variance Asymptotics in Nonstationary VAR's." *Journal of Econometrics*. Forthcoming.

Rajmaira, S. 1995. "India-Pakistani Relations: Reciprocity in Long-Term Perspective." Presented at the annual meeting of the Midwest Political Science Association.

Saikkonen, Pentti. 1991. "Asymptotically Efficient Estimation of Cointegration Relations." *Econometric Theory* 7:1–21.

Schwert, G. W. 1989. "Tests for Unit Roots: A Monte Carlo Investigation." *Journal of Business and Economic Statistics* 7:147–60.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48:1–48.

Sims, Christopher A. 1988. "Bayesian Skepticism on Unit Root Econometrics." *Journal of Economic Dynamics and Control* 12:463–74.

Sims, Christopher A., James H. Stock, and Mark W. Watson. 1990. "Inference in Linear Time Series Models with Some Unit Roots." *Econometrica* 58:113–44.

Sims, Christopher A., and Harald Uhlig. 1991. "Understanding Unit Rooters: A Helicopter Tour." *Econometrica* 59:1591–9.

Singleton, K. J. 1988. "Econometric Issues in the Analysis of Equilibrium Business Cycle Models." *Journal of Monetary Economics* 21:361–86.

Smith, Renée M. 1993. "Error Correction, Attractors, and Cointegration: Substantive and Methodological Issues." *Political Analysis* 4:249–54.

Stock, James H., and Mark W. Watson. 1988. "Testing for Common Trends." *Journal of the American Statistical Association* 83:1097–1107.

Stock, James H., and Mark W. Watson. 1989. "Interpreting the Evidence on Money-Income Causality." *Journal of Econometrics* 40:161–83.

Toda, H., and Peter C. B. Phillips. 1993. "Vector Autoregressions and Causality." *Econometrica* 61:1367–93.

Toda, H., and Taku Yamamoto. 1995. "Statistical Inference in Vector Autoregressions with Possibly Integrated Regressors." *Journal of Econometrics* 66:225–50.

Tsay, Ruey S. 1986. "Time Series Model Specification in the Presence of Outliers." *Journal of the American Statistical Association* 81:132–41.

Williams, John T. 1990. "The Political Manipulation of Macroeconomic Policy." *American Political Science Review* 84:767–96.

Williams, John T. 1993. "What Goes Around Comes Around: Unit Root Tests and Cointegration." *Political Analysis* 4:229–35.

Williams, John T., and Michael McGinnis. 1988. "Sophisticated Reaction in the U.S.-Soviet Arms Race: Evidence of Rational Expectations." *American Journal of Political Science* 32:968–95.

Yamada, H. and H. Toda. 1996. "Inference In Possibly Integrated Vector Autoregressive Models: Finite Sample Evidence." Unpublished manuscript. University of Tsukuba and Osaka University.